



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup>:</b> <b>C12N 15/12, 5/10, 1/21, C07K 14/47, 16/18, C12N 1/21, C07K 14/47, 16/18, C12Q 1/68, G01N 33/50, 33/53, 33/68, A61K 38/17</b>	<b>A2</b>	<b>(11) International Publication Number:</b> <b>WO 98/39448</b> <b>(43) International Publication Date:</b> 11 September 1998 (11.09.98)																														
<b>(21) International Application Number:</b> PCT/US98/04493 <b>(22) International Filing Date:</b> 6 March 1998 (06.03.98) <b>(30) Priority Data:</b> <table border="0"> <tr><td>60/040,162</td><td>7 March 1997 (07.03.97)</td><td>US</td></tr> <tr><td>60/040,333</td><td>7 March 1997 (07.03.97)</td><td>US</td></tr> <tr><td>60/038,621</td><td>7 March 1997 (07.03.97)</td><td>US</td></tr> <tr><td>60/040,161</td><td>7 March 1997 (07.03.97)</td><td>US</td></tr> <tr><td>60/040,626</td><td>7 March 1997 (07.03.97)</td><td>US</td></tr> <tr><td>60/040,334</td><td>7 March 1997 (07.03.97)</td><td>US</td></tr> <tr><td>60/040,336</td><td>7 March 1997 (07.03.97)</td><td>US</td></tr> <tr><td>60/040,163</td><td>7 March 1997 (07.03.97)</td><td>US</td></tr> <tr><td>60/043,580</td><td>11 April 1997 (11.04.97)</td><td>US</td></tr> <tr><td>60/043,568</td><td>11 April 1997 (11.04.97)</td><td>US</td></tr> </table> <p><i>(Continued on the following page)</i></p> <b>(71) Applicant (for all designated States except US):</b> HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US). <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> RUBEN, Steven, M. [US/US]; 18528 Heritage Hills Drive, Olney, MD 20832 (US). ROSEN, Craig, A. [US/US]; 22400 Rolling Hills Road, Laytonsville, MD 20882 (US). FISCHER, Carrie, L. [US/US]; 5810 Hall Street, Burke, VA 22015 (US). SOPENET, Daniel, R. [US/US]; 15050 Stillfield, Place, Centreville, VA 22020 (US). CARTER, Kenneth, C. [US/US]; 11601 Brandy Hall Lane, North Potomac, MD 20878 (US). BEDNARIK, Daniel, P. [US/US]; 8822 Blue Sea Drive, Columbia, MD 21046 (US). ENDRESS, Gregory, A. [US/US]; 9729 Clagett Farm Drive, Potomac, MD 20854 (US). YU, Guo-Liang [CN/US]; 13524 Straw Bale Lane, Darnestown, MD 20878 (US). NI, Jian [CN/US]; 5502 Manorfield Road, Rockville, MD 20853 (US). FENG, Ping [CN/US]; 4 Relda Court, Gaithersburg, MD 20878 (US). YOUNG, Paul, E. [US/US]; 122 Beckwith Street, Gaithersburg, MD 20878 (US). GREENE, John, M. [US/US]; 872 Diamond Drive, Gaithersburg, MD 20878 (US). FERRIE, Ann, M. [US/US]; 13203 L Astoria Hill Court, Germantown,		60/040,162	7 March 1997 (07.03.97)	US	60/040,333	7 March 1997 (07.03.97)	US	60/038,621	7 March 1997 (07.03.97)	US	60/040,161	7 March 1997 (07.03.97)	US	60/040,626	7 March 1997 (07.03.97)	US	60/040,334	7 March 1997 (07.03.97)	US	60/040,336	7 March 1997 (07.03.97)	US	60/040,163	7 March 1997 (07.03.97)	US	60/043,580	11 April 1997 (11.04.97)	US	60/043,568	11 April 1997 (11.04.97)	US	MD 20874 (US). DUAN, Roxanne [US/US]; 4541 Fairfield Drive, Bethesda, MD 20814 (US). HU, Jing-Shan [CN/US]; 1247 Lakeside Drive #3034, Sunnyvale, CA 94086 (US). FLORENCE, Kimberly, A. [US/US]; 12805 Atlantic Avenue, Rockville, MD 20851 (US). OLSEN, Henrik, S. [DK/US]; 182 Kendrick Place #24, Gaithersburg, MD 20878 (US). EBNER, Reinhard [DE/US]; 9906 Shelburne Terrace #316, Gaithersburg, MD 20878 (US). BREWER, Laurie, A. [US/US]; 14920 Mount Nebo Road, Poolesville, MD 20837 (US). MOORE, Paul, A. [GB/US]; Apartment #104, 1908 Holly Ridge Drive, McLean, VA 22102 (US). SHI, Yanggu [CN/US]; 437 West Side Drive, Gaithersburg, MD 20878 (US). LAFLEUR, David, W. [US/US]; 1615 Q Street, N.W. #807, Washington, DC 20009 (US). LI, Yi [CN/US]; 1247 Lakeside Drive #3034, Sunnyvale, CA 94086 (US). ZENG, Zhizhen [CN/US]; 13950 Saddleview Drive, Gaithersburg, MD 20878 (US). KYAW, Hla [BU/US]; 520 Sugarbush Circle, Frederick, MD 21703 (US). <b>(74) Agents:</b> BROOKES, Anders, A. et al.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 10850 (US). <b>(81) Designated States:</b> AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).
60/040,162	7 March 1997 (07.03.97)	US																														
60/040,333	7 March 1997 (07.03.97)	US																														
60/038,621	7 March 1997 (07.03.97)	US																														
60/040,161	7 March 1997 (07.03.97)	US																														
60/040,626	7 March 1997 (07.03.97)	US																														
60/040,334	7 March 1997 (07.03.97)	US																														
60/040,336	7 March 1997 (07.03.97)	US																														
60/040,163	7 March 1997 (07.03.97)	US																														
60/043,580	11 April 1997 (11.04.97)	US																														
60/043,568	11 April 1997 (11.04.97)	US																														
<b>(54) Title:</b> 186 HUMAN SECRETED PROTEINS <b>(57) Abstract</b> <p>The present invention relates to 186 novel human secreted proteins and isolated nucleic acids containing the coding regions of the genes encoding such proteins. Also provided are vectors, host cells, antibodies, and recombinant methods for producing human secreted proteins. The invention further relates to diagnostic and therapeutic methods useful for diagnosing and treating disorders related to these novel human secreted proteins.</p>																																

**Published**

*Without international search report and to be republished upon receipt of that report.*

60/043,314	11 april 1997 (11.04.97)	US	60/047,596	23 May 1997 (23.05.97)	US	60/056,637	22 August 1997 (22.08.97)	US
60/043,569	11 april 1997 (11.04.97)	US	60/047,612	23 May 1997 (23.05.97)	US	60/056,903	22 August 1997 (22.08.97)	US
60/043,311	11 april 1997 (11.04.97)	US	60/047,632	23 May 1997 (23.05.97)	US	60/056,888	22 August 1997 (22.08.97)	US
60/043,671	11 april 1997 (11.04.97)	US	60/047,601	23 May 1997 (23.05.97)	US	60/056,879	22 August 1997 (22.08.97)	US
60/043,674	11 april 1997 (11.04.97)	US	60/047,595	23 May 1997 (23.05.97)	US	60/056,880	22 August 1997 (22.08.97)	US
60/043,669	11 april 1997 (11.04.97)	US	60/047,599	23 May 1997 (23.05.97)	US	60/056,894	22 August 1997 (22.08.97)	US
60/043,312	11 april 1997 (11.04.97)	US	60/047,588	23 May 1997 (23.05.97)	US	60/056,911	22 August 1997 (22.08.97)	US
60/043,313	11 april 1997 (11.04.97)	US	60/047,585	23 May 1997 (23.05.97)	US	60/056,636	22 August 1997 (22.08.97)	US
60/043,672	11 april 1997 (11.04.97)	US	60/047,586	23 May 1997 (23.05.97)	US	60/056,874	22 August 1997 (22.08.97)	US
60/043,578	11 april 1997 (11.04.97)	US	60/047,590	23 May 1997 (23.05.97)	US	60/056,910	22 August 1997 (22.08.97)	US
60/043,576	11 april 1997 (11.04.97)	US	60/047,594	23 May 1997 (23.05.97)	US	60/056,864	22 August 1997 (22.08.97)	US
60/043,670	11 april 1997 (11.04.97)	US	60/047,589	23 May 1997 (23.05.97)	US	60/056,631	22 August 1997 (22.08.97)	US
60/047,600	23 May 1997 (23.05.97)	US	60/047,593	23 May 1997 (23.05.97)	US	60/056,845	22 August 1997 (22.08.97)	US
60/047,615	23 May 1997 (23.05.97)	US	60/047,614	23 May 1997 (23.05.97)	US	60/056,892	22 August 1997 (22.08.97)	US
60/047,597	23 May 1997 (23.05.97)	US	60/047,501	23 May 1997 (23.05.97)	US	60/056,632	22 August 1997 (22.08.97)	US
60/047,502	23 May 1997 (23.05.97)	US	60/048,974	06 June 1997 (06.06.97)	US	60/056,664	22 August 1997 (22.08.97)	US
60/047,633	23 May 1997 (23.05.97)	US	60/048,964	06 June 1997 (06.06.97)	US	60/056,876	22 August 1997 (22.08.97)	US
60/047,583	23 May 1997 (23.05.97)	US	60/049,610	13 June 1997 (13.06.97)	US	60/056,881	22 August 1997 (22.08.97)	US
60/047,617	23 May 1997 (23.05.97)	US	60/051,926	08 July 1997 (08.07.97)	US	60/056,909	22 August 1997 (22.08.97)	US
60/047,618	23 May 1997 (23.05.97)	US	60/052,874	16 July 1997 (16.07.97)	US	60/056,875	22 August 1997 (22.08.97)	US
60/047,503	23 May 1997 (23.05.97)	US	60/055,724	18 August 1997 (18.08.97)	US	60/056,862	22 August 1997 (22.08.97)	US
60/047,592	23 May 1997 (23.05.97)	US	60/056,886	22 August 1997 (22.08.97)	US	60/056,887	22 August 1997 (22.08.97)	US
60/047,581	23 May 1997 (23.05.97)	US	60/056,877	22 August 1997 (22.08.97)	US	60/056,908	22 August 1997 (22.08.97)	US
60/047,584	23 May 1997 (23.05.97)	US	60/056,889	22 August 1997 (22.08.97)	US	60/056,884	22 August 1997 (22.08.97)	US
60/047,500	23 May 1997 (23.05.97)	US	60/056,893	22 August 1997 (22.08.97)	US	60/057,761	05 September 1997 (05.09.97)	US
60/047,587	23 May 1997 (23.05.97)	US	60/056,630	22 August 1997 (22.08.97)	US	60/057,650	05 September 1997 (05.09.97)	US
60/047,492	23 May 1997 (23.05.97)	US	60/056,878	22 August 1997 (22.08.97)	US	60/057,669	05 September 1997 (05.09.97)	US
60/047,598	23 May 1997 (23.05.97)	US	60/056,662	22 August 1997 (22.08.97)	US	60/058,785	12 September 1997 (12.09.97)	US
60/047,613	23 May 1997 (23.05.97)	US	60/056,872	22 August 1997 (22.08.97)	US	60/061,060	02 October 1997 (02.10.97)	US
60/047,582	23 May 1997 (23.05.97)	US	60/056,882	22 August 1997 (22.08.97)	US			

# **FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## 186 Human Secreted Proteins

### *Field of the Invention*

This invention relates to newly identified polynucleotides and the polypeptides encoded by these polynucleotides, uses of such polynucleotides and polypeptides, and their production.

### *Background of the Invention*

Unlike bacterium, which exist as a single compartment surrounded by a membrane, human cells and other eucaryotes are subdivided by membranes into many functionally distinct compartments. Each membrane-bounded compartment, or organelle, contains different proteins essential for the function of the organelle. The cell uses "sorting signals," which are amino acid motifs located within the protein, to target proteins to particular cellular organelles.

One type of sorting signal, called a signal sequence, a signal peptide, or a leader sequence, directs a class of proteins to an organelle called the endoplasmic reticulum (ER). The ER separates the membrane-bounded proteins from all other types of proteins. Once localized to the ER, both groups of proteins can be further directed to another organelle called the Golgi apparatus. Here, the Golgi distributes the proteins to vesicles, including secretory vesicles, the cell membrane, lysosomes, and the other organelles.

Proteins targeted to the ER by a signal sequence can be released into the extracellular space as a secreted protein. For example, vesicles containing secreted proteins can fuse with the cell membrane and release their contents into the extracellular space - a process called exocytosis. Exocytosis can occur constitutively or after receipt of a triggering signal. In the latter case, the proteins are stored in secretory vesicles (or secretory granules) until exocytosis is triggered. Similarly, proteins residing on the cell membrane can also be secreted into the extracellular space by proteolytic cleavage of a "linker" holding the protein to the membrane.

Despite the great progress made in recent years, only a small number of genes encoding human secreted proteins have been identified. These secreted proteins include the commercially valuable human insulin, interferon, Factor VIII, human growth hormone, tissue plasminogen activator, and erythropoietin. Thus, in light of the pervasive role of secreted proteins in human physiology, a need exists for identifying and characterizing novel human secreted proteins and the genes that encode them. This knowledge will allow one to detect, to treat, and to prevent medical disorders by using secreted proteins or the genes that encode them.

### *Summary of the Invention*

The present invention relates to novel polynucleotides and the encoded polypeptides. Moreover, the present invention relates to vectors, host cells, antibodies, and recombinant methods for producing the polypeptides and polynucleotides. Also provided are diagnostic methods for detecting disorders related to the polypeptides, and therapeutic methods for treating such disorders. The invention further relates to screening methods for identifying binding partners of the polypeptides.

### *Detailed Description*

#### Definitions

The following definitions are provided to facilitate understanding of certain terms used throughout this specification.

In the present invention, "isolated" refers to material removed from its original environment (e.g., the natural environment if it is naturally occurring), and thus is altered "by the hand of man" from its natural state. For example, an isolated polynucleotide could be part of a vector or a composition of matter, or could be contained within a cell, and still be "isolated" because that vector, composition of matter, or particular cell is not the original environment of the polynucleotide.

In the present invention, a "secreted" protein refers to those proteins capable of being directed to the ER, secretory vesicles, or the extracellular space as a result of a signal sequence, as well as those proteins released into the extracellular space without necessarily containing a signal sequence. If the secreted protein is released into the extracellular space, the secreted protein can undergo extracellular processing to produce a "mature" protein. Release into the extracellular space can occur by many mechanisms, including exocytosis and proteolytic cleavage.

As used herein, a "polynucleotide" refers to a molecule having a nucleic acid sequence contained in SEQ ID NO:X or the cDNA contained within the clone deposited with the ATCC. For example, the polynucleotide can contain the nucleotide sequence of the full length cDNA sequence, including the 5' and 3' untranslated sequences, the coding region, with or without the signal sequence, the secreted protein coding region, as well as fragments, epitopes, domains, and variants of the nucleic acid sequence. Moreover, as used herein, a "polypeptide" refers to a molecule having the translated amino acid sequence generated from the polynucleotide as broadly defined.

In the present invention, the full length sequence identified as SEQ ID NO:X was often generated by overlapping sequences contained in multiple clones (contig



analysis). A representative clone containing all or most of the sequence for SEQ ID NO:X was deposited with the American Type Culture Collection ("ATCC"). As shown in Table 1, each clone is identified by a cDNA Clone ID (Identifier) and the ATCC Deposit Number. The ATCC is located at 12301 Park Lawn Drive, Rockville, Maryland 20852, USA. The ATCC deposit was made pursuant to the terms of the Budapest Treaty on the international recognition of the deposit of microorganisms for purposes of patent procedure.

A "polynucleotide" of the present invention also includes those polynucleotides capable of hybridizing, under stringent hybridization conditions, to sequences contained in SEQ ID NO:X, the complement thereof, or the cDNA contained within the clone deposited with the ATCC. "Stringent hybridization conditions" refers to an overnight incubation at 42° C in a solution comprising 50% formamide, 5x SSC (750 mM NaCl, 75 mM sodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 µg/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1x SSC at about 65°C.

Also contemplated are nucleic acid molecules that hybridize to the polynucleotides of the present invention at lower stringency hybridization conditions. Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages of formamide result in lowered stringency); salt conditions, or temperature. For example, lower stringency conditions include an overnight incubation at 37°C in a solution comprising 6X SSPE (20X SSPE = 3M NaCl; 0.2M NaH<sub>2</sub>PO<sub>4</sub>; 0.02M EDTA, pH 7.4), 0.5% SDS, 30% formamide, 100 µg/ml salmon sperm blocking DNA; followed by washes at 50°C with 1XSSPE, 0.1% SDS. In addition, to achieve even lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5X SSC).

Note that variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. Typical blocking reagents include Denhardt's reagent, BLOTTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility.

Of course, a polynucleotide which hybridizes only to polyA+ sequences (such as any 3' terminal polyA+ tract of a cDNA shown in the sequence listing), or to a

complementary stretch of T (or U) residues, would not be included in the definition of "polynucleotide," since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-stranded cDNA clone).

5       The polynucleotide of the present invention can be composed of any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. For example, polynucleotides can be composed of single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and  
10   double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, the polynucleotide can be composed of triple-stranded regions comprising RNA or DNA or both RNA and DNA. A polynucleotide may also contain one or more modified bases or DNA or RNA backbones modified for stability  
15   or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications can be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically, or metabolically modified forms.

      The polypeptide of the present invention can be composed of amino acids joined  
20   to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres, and may contain amino acids other than the 20 gene-encoded amino acids. The polypeptides may be modified by either natural processes, such as posttranslational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs,  
25   as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be  
30   branched, for example, as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched, and branched cyclic polypeptides may result from posttranslation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a  
35   nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine,

formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, pegylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. (See, for instance, PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993); POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York, pgs. 1-12 (1983); Seifter et al., Meth Enzymol 182:626-646 (1990); Rattan et al., Ann NY Acad Sci 663:48-62 (1992).)

"SEQ ID NO:X" refers to a polynucleotide sequence while "SEQ ID NO:Y" refers to a polypeptide sequence, both sequences identified by an integer specified in Table 1.

"A polypeptide having biological activity" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. In the case where dose dependency does exist, it need not be identical to that of the polypeptide, but rather substantially similar to the dose-dependence in a given activity as compared to the polypeptide of the present invention (i.e., the candidate polypeptide will exhibit greater activity or not more than about 25-fold less and, preferably, not more than about tenfold less activity, and most preferably, not more than about three-fold less activity relative to the polypeptide of the present invention.)

## **Polynucleotides and Polypeptides of the Invention**

### **FEATURES OF PROTEIN ENCODED BY GENE NO: 1**

This gene is expressed primarily in testes tumor and to a lesser extent in fetal brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer particularly of the testes, and defects of the central nervous system such as seizure and neurodegenerative disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly cancer of the testes and central nervous system,

expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., testes and other reproductive tissue, brain and other tissue of the nervous system, and blood cells, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of testicular cancer and treatment of central nervous system disorders since this gene is primarily expressed in the testes tumor and developing brain.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 2**

This gene is expressed primarily in cancer tissues, such as breast cancer and Wilm's tumor, and to a lesser extent in fetal tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, and/or tumors, particularly, those found in the breast, and developmental abnormalities or disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the glandular tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., mammary tissue, and fetal tissue and, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 314 as residues: Pro-11 to Thr-18, Leu-43 to Pro-50, Gly-64 to Leu-72, and Leu-81 to Lys-86.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of cancers and/or tumors, particularly, those found in the breast since expression is mainly in cancer/tumor tissues. May serve as therapeutic proteins for proliferation/differentiation of fetal tissues.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 3**

This gene is expressed primarily in CD34 depleted buffy coat and to a lesser extent in spleen, chronic lymphocytic leukemia.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: blood disorders or leukemias, diseases of the immune system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for  
10 differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or  
15 cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of blood disorders or  
20 leukemias, diseases of the immune system since expression is in tissues related to immune function.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 4**

This gene is expressed primarily in CD34 depleted buffy coat.

25 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: blood disorders or lymphocytic diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification  
30 of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual  
35 having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of blood disorders since expression is in tissues related to immune function.

## 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 5

This gene is expressed primarily in CD34 depleted buffy coat.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: blood or immune  
10 diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and cancerous  
15 and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 317 as residues:  
20 Pro-13 to Lys-21.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of blood disorders since expression is in tissues related to immune function.

## 25 FEATURES OF PROTEIN ENCODED BY GENE NO: 6

This gene is expressed primarily in CD34 depleted buffy coat.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: blood or immune  
30 diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., and blood cells, and  
35 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level

in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 318 as residues: Lys-31 to Lys-39.

5 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of blood diseases since it is expressed in tissues related to immune function.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 7**

10 This gene is expressed primarily in CD34 depleted buffy coat and to a lesser extent in pineal gland.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diseases of the immune system and brain associated diseases. Similarly, polypeptides and antibodies directed to  
15 these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and pineal gland, and cancerous and wounded tissues) or  
20 bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides  
25 corresponding to this gene are useful for treatment/diagnosis of blood disorders, immune diseases or brain associated diseases (specifically of the pineal gland) since expression is in tissues related to immune function.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 8**

30 The translation product of this gene shares sequence homology with an organic cation transporter which is thought to be important in organic cation uptake in the kidney and liver. (See Accession No. 2343059.) Preferred polypeptide fragments comprise the amino acid sequence ITIAIQMICLVNXELYPTFVRNXGVMVCSSLCDIGGIITP FIVFRLREVWQALPLILFAVLGLLAAGVTL LLPETKGVALPETMKDAENLGRKAKPKENTTYLK  
35 VQTSEPSGT (SEQ ID NO: 615) or TMKDAENLGRKAKPKENT (SEQ ID NO: 616) as well as N-terminal and C-terminal deletions of these fragments. Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hepatic and renal diseases where drug elimination/cation exchange (organic cation uptake) in the liver and kidney are problematic. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hepatic or renal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., kidney and liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 320 as residues: Asn-64 to Asn-74, and Gln-81 to Gly-87.

The tissue distribution and homology to organic cation transporter indicate that polynucleotides and polypeptides corresponding to this gene are useful as a polyspecific transporter that is important for drug elimination in the liver (and possibly kidney) since expression is found in the liver.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 9**

This gene is expressed primarily in eosinophil induced with IL-5 and to a lesser extent in fetal liver and spleen. This gene also maps to chromosome 15, and therefore can be used in linkage analysis as a marker for chromosome 15.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diseases of the immune system, particularly allergies or asthma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, liver, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the



standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating/diagnosis of diseases involving eosinophil reactions since expression seems to be concentrated in eosinophils and other tissues involved in immunity such as the liver and spleen.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 10**

This gene is expressed primarily in tissues of hematopoietic lineage and to a lesser extent in Hodgkins lymphoma. Any frame shifts in this sequence can easily be clarified using known molecular biology techniques.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, and immune deficiency or dysfunction. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., hematopoietic cells, lymphoid and reticuloendothelial tissues, and cancerous tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/ diagnosis for lymphomas or immune dysfunction or as a therapeutic protein useful in immune modulation based on expression in anergic T-cells and lymphomas.

#### **30 FEATURES OF PROTEIN ENCODED BY GENE NO: 11**

This gene is expressed primarily in neutrophils and to a lesser extent in activated lymphoid cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the cell type present in a biological sample and for diagnosis of diseases and conditions: inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders

of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 323 as residues: Glu-40 to Lys-46.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for modulation of an immune reaction or as a growth factor for the differentiation or proliferation of neutrophils for the treatment of neutropenia.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 12**

This gene is expressed primarily in brain and to a lesser extent in activated T-cells. It is likely that the open reading frame containing the predicted signal peptide continues in the 5' direction. Preferred polypeptide fragments comprise the amino acid sequence PRVRNSPEDLGLSLTGDCKL (SEQ ID NO:617).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurodegenerative disorders including ischemic shock, alzheimers and cognitive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and brain, and other tissue of the nervous system and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 324 as residues: Ser-5 to Glu-14, Ile-21 to Pro-35, Ser-65 to Asp-81, Cys-89 to Val-96, Lys-136 to Ser-145, Ile-152 to Met-169, and Arg-189 to Lys-196.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnostic/treatment for cancers of the given tissue or in the treatment of neurological disorders of the CNS.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 13**

This gene was also recently cloned by other groups, naming this calcium-activated potassium channel gene, hKCa4. (See Accession No. AF033021, see also, Accession No. 2584866.) This gene is mapped to human chromosome 19q13.2. A second signal sequence likely exists upstream from the predicted signal sequence as described in Table 1. Preferred polypeptide fragments comprise: QADDLQATVAALCVLRGGGPWAG SWLSPKTPGAMGGDLVLGLGALRRRKRL (SEQ NO: 618); or EQEKSLAGWALVLAXXGIGL MVLHAEMLWFGGCSAVNATGHLSDTLWLIPITFLTIGYGDVVPGTMWGKIVCLCTGVMGVCC TALLVAVVARKLEFNKAEKHVHNFMMDIQYTKEMKESAAARVLQEAWMFYKHTRRKESHAAR XHQRXLLAAINAFRQVRLKHRKLREQVNSMVDISKMHMILYDLQNLSSSHRALEKQIDTLAG KLDALTELLSTALGPRQLPEPSQQSK (SEQ ID NO: 619), as well as N-terminal and C-terminal deletions. Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in breast lymph node and T-cells, and to a lesser extent in placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hematologic and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., lymphoid tissue, blood cells and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 325 as residues: Arg-13 to Lys-23.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment/diagnosis of hematologic and diseases involving immune modulation based on distribution in the lymph node and T-cells.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 14**

This gene was recently cloned by another group, calling it PAPS synthase. (See Accession No. e1204135.) Preferred polypeptide fragments comprise the amino acid sequence YQAHVSRNKRQGVVGTRGGFRGCTVWLTGLSGAGK (SEQ ID NO: 620).

- 5 Also preferred are the polynucleotide fragments encoding this polypeptide fragment.

It has been discovered that this gene is expressed primarily in benign prostate hyperplasia, Human Umbilical Vein Endothelial Cells and to a lesser extent in smooth muscle and Human endometrial stromal cells-treated with estradiol.

- 10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: inflammation, ischemia, and restenosis, based on endothelial cell and smooth muscle cell expression, and prostate diseases such as benign prostate hyperplasia or prostate cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing
- 15 immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the prostate or vessels of the circulatory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., prostate, endothelial cells, smooth muscle, and endometrium, and cancerous and wounded tissues) or bodily
- 20 fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 326 as residues: Arg-21 to Asp-26, Lys-35 to Lys-44,
- 25 Glu-49 to Asn-58.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating/diagnosing diseases or conditions where the endothelial cell lining of the veins and arteries of underlying smooth muscle are involved.

30

**FEATURES OF PROTEIN ENCODED BY GENE NO: 15**

This gene is expressed primarily in human 6 week embryo and to a lesser extent in placenta.

- 35 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: developmental anomalies or fetal deficiencies. Similarly, polypeptides and antibodies directed to these

polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly developmental in nature, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic tissue, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 327 as residues Lys-50 to Glu-57.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection of developmental abnormalities.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 16**

This gene is expressed primarily in kidney and amygdala and to a lesser extent in fetal tissues. This gene is mapped to chromosome 14, and therefore is useful in linkage analysis as a marker for chromosome 14.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) present in a biological sample and for diagnosis of diseases and conditions: kidney diseases, neurological disorders and developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s). For a number of disorders of the above tissues, particularly of the renal system or developing fetal tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., kidney, amygdala, and fetal tissues, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment or diagnosis of conditions affecting the brain, kidneys and fetal development.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 17**

This gene is expressed primarily in ovarian cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: solid tumors similar to ovarian cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., ovarian and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 329 as residues Ser-51 to Val-56.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of solid tumors of the reproductive system such as ovarian cancer.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 18**

This gene is expressed primarily in brain medulloblastoma. Preferred polypeptide fragments comprise the amino acid sequence: IRHEQHPNFSLEMHSGSSLLFLPQL ILILPVCAHLHEELNC (SEQ ID NO: 643) and SFFISEEKGHLLQAERHPWVAGALVGVSGLTLTTCSGPTEKPATKNYFLKRLQEMHIRAN (SEQ ID NO: 644), as well as N-terminal and C-terminal deletions. Also preferred are polynucleotide fragments encoding these polypeptide fragments.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors particularly of the CNS or. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the Central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene

expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating medulloblastoma or similar tumors.

5

#### **FEATURES OF PROTEIN ENCODED BY GENE NO.: 19**

This gene is expressed primarily in adipocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: obesity. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the adipose tissues expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., adipocytes and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating obesity by regulating the function and number of adipocytes

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 20**

This gene is expressed primarily in B cell lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, of the immune system with an emphasis on B cell lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the tumors of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of B cell derived tumors based on its expression in b cell lymphomas

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 21**

This gene is expressed primarily in immune cells and to a lesser extent in fetal tissues

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: inflammatory diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., cells of the immune system, and fetal tissues, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:333 as residues Asp-10 to Pro-19, Ser-74 to Tyr-79, Glu-95 to Lys-110.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of diseases involving alterations in T cell activity.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 22**

It has been discovered that this gene is expressed primarily in ovarian tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors particularly of the ovary. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of tumors of the reproductive organs. expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., ovarian



and other reproductive tissue and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 334 as residues: Leu-22 to Gln-27.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of ovarian tumors as it has only been identified in ovarian tumors.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 23

It has been discovered that this gene is expressed primarily in fetal tissues and to a lesser extent in osteoclastoma cell line

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: osteoporosis or arthritis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone cells, and fetal tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of conditions of abnormal bone remodeling due to enhanced activity of osteoclasts. This may be useful as a specific marker for malignancies derived from osteoclasts or their precursors.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 24

The translation product of this gene shares sequence homology with a periplasmic ribonuclease which is thought to be important in degrading extracellular polynucleotides

It has been discovered that this gene is expressed primarily in serum treated smooth muscle cells

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: vascular disease such as restenosis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vasculature expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., smooth muscle, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 336 as residues: Gln-30 to Lys-36, and Pro-41 to Arg-48.

The tissue distribution and homology to ribonucleases indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment of pathological conditions of smooth muscle associated with bacterial or viral infiltration

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 25**

This gene is expressed primarily in Early Stage Human Brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: human brain development and related diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the human brain development and related diseases, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to this gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases affecting human brain development and related diseases.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 26**

It has been discovered that this gene is expressed primarily in human brain tissue.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: human brain diseases and other diseases related to brain diseases, which may be caused by brain diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in  
10 providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the human brain diseases, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum,  
15 plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to the gene indicate that polynucleotides  
20 and polypeptides corresponding to this gene are useful for diagnosis and treatment of human brain diseases and other diseases related.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 27**

It has been discovered that this gene is expressed primarily in Anergic T-cells.

25 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune diseases, inflammatory diseases and diseases related to T lymph cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological  
30 probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune diseases, inflammatory diseases and diseases related to T lymph cells, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and cancerous and wounded tissues) or bodily fluids (e.g.,  
35 serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene

expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to the gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for immune diseases,  
5 inflammatory diseases and diseases related to T lymph cells.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 28

The translation product of this gene shares sequence homology with *Shigella flexneri* positive transcriptional regulator CriR (criR) gene which is thought to be  
10 important in regulation of gene expression.

This gene is expressed primarily in human synovial sarcoma and normal human brain tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a  
15 biological sample and for diagnosis of diseases and conditions: human brain diseases particularly sarcomas of the synovium. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the human brain and synovium and other related human  
20 brain diseases, expression of this gene at significantly higher or lower levels may be routinely detected in certain (e.g., synovial tissue, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,  
25 the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of human synovial sarcoma and other related human brain diseases.  
30

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 29

This gene is expressed in bone marrow, infant brain, fetal liver and spleen, prostate and to a lesser extent in pineal gland, adipose tissue, kidney, adrenal gland, umbilical vein endothelial cells, and T cells.

35 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diseases related to bone marrow or

hematoplastic tissues, prostate, kidney, adrenal gland, and cardiovascular tissue or organs. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the diseases related to hematoplastic tissues, immune system, prostate, kidney, adrenal gland, and cardiovascular tissue or organs, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone marrow, hematopoietic cells, pineal gland, adipose tissue, kidney, adrenal gland, endothelial cells, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to the gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases related to hematoplastic tissues, immune system, prostate, kidney, adrenal gland, and cardiovascular tissue or organs.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 30**

This gene is expressed primarily in meningea and to a lesser extent in breast and adult brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: Diseases of the meningea and related brain diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the meningea and related brain diseases, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., meningea, mammary tissue, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases of the meningea and related brain diseases.

#### 5    **FEATURES OF PROTEIN ENCODED BY GENE NO: 31**

This gene is expressed in meningea, fetal spleen, osteoblast and to a lesser extent in activated T-cells, endometrial stromal cells, fetal lung, HL-60, thymus, testis and endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as  
10 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: meningeal disease, osteoporosis, immune diseases, and hematoplastic diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for identification of the tissue(s) or cell type(s). For a number of disorders of the  
15 above tissues or cells, particularly of the meningeal diseases, osteoporosis, immune diseases, and hematoplastic diseases, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, endometrium, lung, thymus, testis, and endothelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal  
20 fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of  
25 meningeal, osteoporosis, immune diseases, hematoplastic diseases, testis diseases and lung diseases.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 32**

This gene is expressed primarily in human thymus and to a much lesser extent  
30 in infant brain, T-cells, smooth muscle, endothelial cells, bone marrow, human ovarian tumor and keratinocytes testes, osteoclastoma, breast, and tonsils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: Diseases involving the  
35 thymus, particularly thymic cancer and diseases involving T-cell maturation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a

number of disorders of the above tissues or cells, particularly of the thymus, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., thymus, brain, and other tissue of the nervous system, blood cells, bone marrow, ovaries, and testes, and other reproductive tissue, mammary tissue, tonsils, melanocytes and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- 10       The tissue distribution and homology to gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases of the thymus particularly thymic cancer and diseases involving T-cell maturation.

15   **FEATURES OF PROTEIN ENCODED BY GENE NO: 33**

      This gene is expressed primarily in human tonsils, and placenta, and to a lesser extent in adipocytes, melanocyte, and infant brain.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: inflammatory diseases, immune diseases, and obesity. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the inflammatory diseases, immune diseases, and obesity, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., tonsils, placenta, adipocytes, melanocytes, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

      The tissue distribution and homology to this gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases such as inflammation, immune diseases, and obesity.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 34**

This gene is expressed in activated T cells, and to a lesser extent in pituitary, testis, and breast lymph node.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diseases relating to T cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the disorders of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., pituitary, testes and other reproductive tissue, mammary tissue, and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of immune disorders.

**20 FEATURES OF PROTEIN ENCODED BY GENE NO: 35**

This gene is expressed primarily in infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the diseases relating to neurological disorders, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain, and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neurological disorders.



**FEATURES OF PROTEIN ENCODED BY GENE NO: 36**

This gene is expressed primarily in infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as  
5 reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions: neurological disorders.  
Similarly, polypeptides and antibodies directed to these polypeptides are useful in  
providing immunological probes for differential identification of the tissue(s) or cell  
type(s). For a number of disorders of the above tissues or cells, particularly of the  
10 diseases relating to neurological disorders, expression of this gene at significantly  
higher or lower levels may be routinely detected in certain tissues and cell types (e.g.,  
brain and other tissue of the nervous system, and cancerous and wounded tissues) or  
bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another  
tissue or cell sample taken from an individual having such a disorder, relative to the  
15 standard gene expression level, i.e., the expression level in healthy tissue or bodily  
fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides  
corresponding to this gene are useful for diagnosis and treatment of neurological  
disorders.

20

**FEATURES OF PROTEIN ENCODED BY GENE NO: 37**

This gene is expressed primarily in human ovary.

Therefore, polynucleotides and polypeptides of the invention are useful as  
reagents for differential identification of the tissue(s) or cell type(s) present in a  
25 biological sample and for diagnosis of diseases and conditions: ovarian cancer.  
Similarly, polypeptides and antibodies directed to these polypeptides are useful in  
providing immunological probes for differential identification of the tissue(s) or cell  
type(s). For a number of disorders of the above tissues or cells, particularly of the  
ovarian disorders such as those involving germ cells, ovarian follicles, stromal cells,  
30 expression of this gene at significantly higher or lower levels may be routinely detected  
in certain tissues and cell types (e.g., ovary and other reproductive tissue, and  
cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial  
fluid or spinal fluid) or another tissue or cell sample taken from an individual having  
such a disorder, relative to the standard gene expression level, i.e., the expression level  
35 in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides  
corresponding to this gene are useful for diagnosis and treatment of ovariothy.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 38**

This gene is expressed primarily in lymph node breast cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as  
5 reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions: breast cancer. Similarly,  
polypeptides and antibodies directed to these polypeptides are useful in providing  
immunological probes for differential identification of the tissue(s) or cell type(s). For a  
number of disorders of the above tissues or cells, particularly of the breast cancer,  
10 expression of this gene at significantly higher or lower levels may be routinely detected  
in certain tissues and cell types (e.g., mammary tissue and lymphoid tissue, and  
cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial  
fluid or spinal fluid) or another tissue or cell sample taken from an individual having  
such a disorder, relative to the standard gene expression level, i.e., the expression level  
15 in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides  
corresponding to this gene are useful for used as a diagnostic marker for breast cancer.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 39**

20 This gene is expressed primarily in brain and to a lesser extent in other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as  
reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions: neuronal disorders such  
as trauma, brain degeneration, and brain tumor. Similarly, polypeptides and antibodies  
25 directed to these polypeptides are useful in providing immunological probes for  
differential identification of the tissue(s) or cell type(s). For a number of disorders of  
the above tissues or cells, particularly of the brain, expression of this gene at  
significantly higher or lower levels may be routinely detected in certain tissues and cell  
types (e.g., brain and other tissue of the nervous system, and cancerous and wounded  
30 tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or  
another tissue or cell sample taken from an individual having such a disorder, relative to  
the standard gene expression level, i.e., the expression level in healthy tissue or bodily  
fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides  
35 corresponding to this gene are useful for diagnosis and therapeutic treatment of  
neuronal disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 40**

5 This gene is expressed in early stage human embryo, adrenal gland tumor, and immune tissues such as fetal liver, fetal spleen, T-cell, and myeloid progenitor cell line and to a lesser extent in ovary, colon cancer, and a few other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumorigenesis including  
10 adrenal gland tumor, colon cancer and various other tumors, developmental and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cancer tissues, early stage human tissues, and immune system,  
15 expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, spleen, blood cells, bone marrow, ovary and other reproductive tissue, and colon, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard  
20 gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and therapeutic treatment of immune and developmental disorders, and tumorigenesis.

25

**FEATURES OF PROTEIN ENCODED BY GENE NO: 41**

This gene is expressed primarily in fetal lung, endothelial cells, liver, thymus and a few other immune tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as  
30 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune disorders such as immune deficiency and autoimmune diseases, pulmonary diseases, liver diseases, and tumor matasis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification  
35 of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetal lung, liver, endothelial cells, and immune tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain

tissues and cell types (e.g., lung, endothelial cells, liver, thymus, and other tissue of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis of immune disorders and pulmonary and hepatic diseases. Its promoter may also be used for immune system and lung-specific gene therapies. The expression of this gene in endothelial cells indicates that it may also involve in angiogenesis which therefore may play role in tumor matasis.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 42**

This gene is expressed primarily in liver, thyroid, parathyroid and to a lesser extent in fetal lung, stomach and early embryos.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: metabolic regulation, obesity, hepatic failure, hepatocellular tumors or thyroiditis and thyroid tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive/endocrine system expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, thyroid, parathyroid, lung, stomach, and embryonic tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and the extracellular locations indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of digestive/endocrine disorders, including metabolic regulation, hepatic failure, malabsorption, gastritis and neoplasms.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 43**

This gene is expressed primarily in Schizophrenic adult brain, pituitary, front cortex, hypothalamus and to a lesser extent in retina, adipose and stomach cancer and placenta.

- 5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: schizophrenia and other neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification
- 10 of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nerve system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., retinal tissue, adipose, stomach, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or
- 15 cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful in treatment/detection of disorders in the nerve
- 20 system, including schizophrenia, neurodegeneration, and neoplasia. Additionally, a secreted protein in brain may serve as an endocrine.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 44**

- The translation product of this gene shares sequence homology with GTP
- 25 binding proteins which are thought to be important in signal transduction and protein transport.

This gene is expressed primarily in umbilical vein and microvascular endothelial cells, GM-CSF treated macrophage, anergic T cells, osteoblast, osteoclast, CD34+ cells and to a lesser extent in gall bladder.

- 30 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: bone formation and growth, osteonecrosis, osteoporosis, angiogenesis and/or hematopoiesis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing
- 35 immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal and hematopoiesis systems, expression of this gene at significantly higher or lower levels

may be routinely detected in certain tissues and cell types (e.g., endothelial cells, blood cells, bone, and gall bladder, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to GTP binding proteins indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment/detection of bone formation and growth, osteonecrosis, osteoporosis, and/or hematopoiesis because its involvement in the growth signaling or angiogenesis.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 45**

The translation product of this gene shares sequence homology with signal sequence receptor gamma subunit which is thought to be important in protein translocation on endoplasmic reticulum.

This gene is expressed primarily in adrenal gland, salivary gland, prostate, and to a lesser extent in endothelial cells and smooth muscle.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: protein secretion. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the secretory organs, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., adrenal gland, salivary gland, prostate, endothelial cells, and smooth muscle, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to SSR gamma subunit indicate that polynucleotides and polypeptides corresponding to this gene are useful for endocrine disorders, prostate cancer, xerostomia or sialorrhea.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 46**

This gene is expressed primarily in osteoclastoma cells and to a lesser extent in melanocyte, amygdala, brain, and stomach.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: ossification, osteoporosis, fracture, osteonecrosis, osteosarcoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., melanocytes, amygdala, brain and other tissue of the nervous system, and stomach, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful in intervention of ossification, osteoporosis, fracture, osteonecrosis and osteosarcoma.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 48**

The translation product of this gene shares sequence homology with proline rich proteins which is thought to be important in protein-protein interaction.

This gene is expressed primarily in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurological and psychological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nerve system and endocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to proline-rich proteins indicate that polynucleotides and polypeptides corresponding to this gene are useful in intervention

and detection of neurological diseases, including trauma, neoplasia, degenerative or metabolic conditions in the central nerve system. Additionally, the gene product may be a secreted by the brain as an endocrine.

#### 5 **FEATURES OF PROTEIN ENCODED BY GENE NO: 49**

The translation product of this gene shares sequence homology with the AOCB gene from *Aspergillus nidulans* which is important in asexual development.

This gene is expressed primarily in infant brain and to a lesser extent in the developing embryo, trachea tumors, B-cell lymphoma and synovial sarcoma.

10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurodegenerative diseases, leukemia and sarcoma's. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential  
15 identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic tissue, blood cells, trachea, and synovial tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or  
20 spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in infant brain and sarcoma's and homology to a gene involved in a key step of eukaryotic development (fungal spore formation) indicates  
25 that the protein product of this clone could play a role in neurological diseases such as schizophrenia, particularly in infants. The existence of the gene in a B-cell lymphoma indicates the gene may be used in the treatment and detection of leukemia.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 50**

30 This gene is expressed primarily in fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: pulmonary disorders including lung cancer. Similarly, polypeptides and antibodies directed to these  
35 polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the pulmonary system, expression of this gene at significantly higher or



lower levels may be routinely detected in certain tissues and cell types (e.g., lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene only in fetal lung indicates that it plays a key role in development of the pulmonary system. This would suggest that misregulation of the expression of this protein product in the adult could lead to lymphoma or sarcoma formation, particularly in the lung. It may also be involved in predisposition to certain pulmonary defects such as pulmonary edema and embolism, bronchitis and cystic fibrosis.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 51**

This gene is expressed primarily in hematopoietic cell types and fetal cells and to a lesser extent in all tissue types.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects in the immune system and hematopoiesis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., hematopoietic cells, and fetal tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene predominantly in hematopoietic cells and in the developing embryo indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection and treatment of lymphomas and disease states affecting the immune system or hematopoiesis disorders such as leukemia, AIDS, arthritis and asthma..

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 52**

This gene is expressed primarily in prostate and to a lesser extent in fetal spleen, fetal liver, infant brain and T cell leukemias.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: prostate disorders, prostate cancer, leukemia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, and/or prostate gland expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., thymus, spleen, liver, brain and other tissue of the nervous system, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene in prostate indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection or treatment of prostate disorders or prostate cancer. Its distribution in fetal liver and fetal spleen indicates it may play a role in the immune system and its misregulation could lead to immune disorders such as leukemia, arthritis and asthma.

20

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 53**

The translation product of this gene shares sequence homology with dynein.

This gene is expressed primarily in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neuro-degenerative diseases of the brain. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly neuro-degenerative diseases expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The predominant tissue distribution in the brain and homology to dynein, a microtubule motor protein involved in the positioning of cellular organelles and molecules indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection/treatment of neurodegenerative diseases, such as Alzheimers, 5   Huntigtons, Parkinsons diseases and shizophrenia.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 54**

The translation product of this gene shares sequence homology with ubiquitin-conjugation protein, an enzyme which is thought to be important in the processing of 10   the Huntingtons Disease causing gene.

This gene is expressed primarily in brain and to a lesser extent in activated macrophages.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a 15   biological sample and for diagnosis of diseases and conditions: neurodegenerative disease states including Huntington's disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of brain tissues. For a number of disorders of the above tissues or cells, particularly of the neurological systems expression of this gene at 20   significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level 25   in healthy tissue or bodily fluid from an individual not having the disorder.

The predominant tissue distribution of this gene in the brain and its homology to a Huntington interacting protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for the regulation of the expression of the 30   Huntington disease gene and other neurodegenerative diseases including spinocerebellar ataxia types I and III, dentatorubropallidoluysian and spinal bulbar muscular atrophy. In addition, the existence of elevated levels of free ubiquitin pools in Alzheimer's disease, Parkinson's disease and amyotrophic lateral sclerosis indicates that the ubiquitin pathway of protein degradation plays a role in these disease states. Thus, considering the gene described here is homologous to a ubiquitin-conjugation 35   protein it may play a general role in neurodegenerative conditions.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 56**

This gene is expressed primarily in T-cells (anergic T-cells, resting T-Cells, apoptotic T-cells) and lymph node (breast), as well as brain (hypothalamus, hippocampus, pituitary, infant brain, early-stage brain).

- 5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune (e.g. immunodeficiencies, autoimmunities, inflammation, leukemias & lymphomas) and neurological (e.g. Alzheimer's disease, dementia, schizophrenia) disorders. Similarly,
- 10 polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous, hematopoietic and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood
- 15 cells, lymphoid tissue, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.
- 20 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful in the intervention or detection of pathologies associated with the hematopoietic and immune systems, such as anemias (leukemias). In addition, the expression in brain (including fetal) might suggest a role in developmental brain defects, neuro-degenerative diseases or behavioral abnormalities
- 25 (e.g. schizophrenia, Alzheimer's, dementia, depression, etc.).

**FEATURES OF PROTEIN ENCODED BY GENE NO: 57**

- This gene is expressed primarily in lung, and to a lesser extent in a variety of other hematological cell types (e.g. Raji cells, bone marrow cell line, activated
- 30 monocytes).

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: pulmonary and/or hematological disfunction. Similarly, polypeptides and antibodies directed to these
- 35 polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vasculo-pulmonary and hematopoietic systems, expression of this

gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., lung and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful in the intervention and detection of pathologies associated with the vasculo-pulmonary system. In addition the expression of this gene in a variety of leukocytic cell types and a bone marrow cell line might suggest a role in hematopoietic and immune system disorders, such as leukemias & lymphomas, inflammation, immunodeficiencies and autoimmunities.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 58**

The translation product of this gene shares sequence homology with adenylate kinase isozyme 3 (gil163528 GTP:AMP phosphotransferase (EC 2.7.4.10) [Bos taurus]), which is thought to be important in catalyzing the phosphorylation of AMP to ADP in the presence of ATP or inorganic triphosphate.

This gene is expressed primarily in fetal liver, heart and placenta, and to a lesser extent in many other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hepatic, cardiovascular or reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hepatic, cardiovascular and reproductive systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, heart, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of conditions related to hepatic function and pathogenesis, in particular, those dealing with liver development and the differentiation of hepatocyte progenitor cells.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 59**

This gene is expressed primarily in CD34 positive cells (Cord Blood).

Therefore, polynucleotides and polypeptides of the invention are useful as  
5 reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions: hematopoietic  
differentiation and immune disorders. Similarly, polypeptides and antibodies directed to  
these polypeptides are useful in providing immunological probes for differential  
10 identification of the tissue(s) or cell type(s). For a number of disorders of the above  
tissues or cells, particularly of hematopoietic and immune systems, expression of this  
gene at significantly higher or lower levels may be routinely detected in certain tissues  
and cell types (e.g., hematopoietic cells, and blood cells, and cancerous and wounded  
tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or  
15 another tissue or cell sample taken from an individual having such a disorder, relative to  
the standard gene expression level, i.e., the expression level in healthy tissue or bodily  
fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides  
corresponding to this gene are useful in the detection and treatment of conditions  
associated with CD34-positive cells, and therefore as a marker for cell differentiation in  
20 hematopoiesis, as well as immunological disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 60**

The translation product of the predicted open reading frame of this contig has  
sequence identity to the murine gene designated Insulin-Like Growth Factor-Binding  
25 Protein (IGFBP)-1 as described by Lee and colleagues (Hepatology 19 (3), 656-665  
(1994)).

This gene is expressed exclusively in hemangiopericytoma.

Therefore, polynucleotides and polypeptides of the invention are useful as  
reagents for differential identification of the tissue(s) or cell type(s) present in a  
30 biological sample and for diagnosis of hemangiopericytoma and other pericyte or  
endothelial cell proliferative disorders. Similarly, polypeptides and antibodies directed  
to these polypeptides are useful in providing immunological probes for differential  
identification of the tissue(s) or cell type(s). For a number of disorders of the above  
tissues or cells, particularly of the circulatory and immune systems, expression of this  
35 gene at significantly higher or lower levels may routinely be detected in certain tissues  
and cell types (e.g., pericyte or endothelial cells, and liver, and cancerous and wounded  
tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or

another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Polynucleotides and polypeptides corresponding to this gene are useful as cell growth regulators since IGFBP-1-like molecules function as modulators of insulin-like growth factor activity. In addition, since IGFBP-1 is expressed at high levels following hepatectomy and during fetal liver development, polynucleotides of the present invention may also be used for the diagnosis of developmental disorders. Further, polypeptides of the present invention may be used therapeutically to treat developmental liver disorders as well as to regulate hepatocyte and supporting cell growth following hepatectomy or to treat liver disorders.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of hemangiopericytoma and liver disorders.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 61**

This gene is expressed primarily in schizophrenic frontal cortex.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: nervous system and cognitive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the frontal cortex and CNS expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, treatment and diagnosis of frontal cortex, neuro-degenerative and CNS disorders

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 62**

This gene is expressed primarily in human adrenal gland tumor, and to a lesser extent in human kidney, medulla and adult pulmonary tissue.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: metabolic, endocrine disorders. Similarly, polypeptides and antibodies directed to these polypeptides are
- 5 useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine and nervous system disorders and neoplasia, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., adrenal gland, kidney, brain and other tissue of the nervous system,
- 10 pulmonary tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.
- 15 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, treatment and diagnosis of neurological and endocrine disorders including neoplasia.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 63**

- 20 This gene is expressed primarily in human adipocytes, and to a lesser extent in spleen, 12-week old human, and testes.
- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune, metabolic and
- 25 growth disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., adipocytes,
- 30 spleen, and testes and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.
- 35 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis and treatment of immune, developmental and metabolic disorders.



**FEATURES OF PROTEIN ENCODED BY GENE NO: 64**

One translated product of this clone is homologous to the mouse zinc finger protein PZF. (See Accession No. 453376; see also Gene 152 (2), 233-238 (1995).) Preferred polypeptide fragments correspond to the highly conserved domains shared between mouse and man. For example, preferred polypeptide fragments comprise the amino acid sequence: LQCEICGFTCRQKASLNWHMKKHDADSFYQFSCNICGKKFEKKDSVVAHKAKSH PEV (SEQ ID NO: 621); ITSTDILGTNPESLTQPSD (SEQ ID NO: 622); NSTSGECLLLEAEGM SKSY (SEQ ID NO: 623); CSGTERVSLMADGKIFVGS GSGGTEGLVMNSDILGATTEVLIEDSD SAGP (SEQ ID NO: 624); IQYVRCEMEGCGTVLAHPRYLQHIIKYQHLLKKKYVCPHPSCGRLF RLQKQLLRHAKHHT (SEQ ID NO: 625); DQRDYICEYCARAFKSSHNLAVHRMIHTGEK (SEQ ID NO: 626); RSSRTSVSRHRDTENTRSSRSKTGSLQLICKSEPNTDQLDY (SEQ ID NO: 627); PFKDDPRDETYKPHLERETPKPRRKSG (SEQ ID NO: 630); QYVRCEMEGCGTVLAHPRYLQ HHIKYQHLLKKKYVCPHPSCGRLFRLQKQLLRHAKHHTD (SEQ ID NO: 629); or residues 151-182 of QRDYICEYCARAFKSSHNLAVHRMIHTGEKHY (SEQ ID NO: 628). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in Rhabdomyosarcoma, melanocyte and colon cancer tissue and to a lesser extent in smooth muscle, pancreatic tumor, and apoptotic T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to,. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hemopoetic, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., striated muscle, melanocytes, colon, smooth muscle, pancreas, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis and treatment of cancer and hemopoetic disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 65**

This gene is expressed primarily in human adipose and salivary gland tissue and to a lesser extent in human bone marrow and fetal kidney.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: metabolic and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the metabolic and hemopoetic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., adipose, salivary gland, bone marrow, and kidney, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis of metabolic and immune disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 66**

This translated product of this gene was recently identified as oxytocinase splice variant 1. (See Accession Nos. 2209276 and d1010078.) Preferred polypeptide fragments comprise the amino acid sequence: EMFDSL SYFKGSSLLMLKTYLSEDVFQHAVVLYLHN HSYASIQSDDLWDSFNEVTNQILDVKRMMKWTWLQKGFPLVTQKKGKELFIQQRFFLNMK PEIQPSDTRYM (SEQ ID NO: 631). Also preferred are polynucleotide fragments encoding this polypeptide fragment.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 67**

This gene is expressed primarily in hemopoetic cells, particularly apoptotic T-cells, and to lesser extent in primary dendritic cells and adipose tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of apoptotic T-cells, primary dendritic cells, and adipose tissue present in a biological sample and for diagnosis of diseases and conditions: hemopoetic diseases including cancer and general immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell

type(s). For a number of disorders of the above tissues or cells, particularly of the oral and intestinal mucosa as well as hemopoetic and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., hematopoietic cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of diseases of the immune system, including cancer, hemopoetic and infectious diseases.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 68**

This gene is expressed primarily in kidney cortex and to a lesser extent in infant brain, heart, uterus, and blood.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of kidney tissue present in a biological sample and for diagnosis of diseases and conditions: soft tissue cancer, inflammation, kidney fibrosis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and endocrines systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., kidney, brain, and other nervous tissue, heart, uterus, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study and treatment of cancer and fibroses.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 69**

The translation product of this gene shares strong sequence homology with vertebrate and invertebrate protein tyrosine phosphatases.

This gene is expressed primarily in endometrial tumors, melanocytes, myeloid progenitors and to a lesser extent in infant brain, adipocytes, and several hematopoietic stem cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of transformed hematopoietic and epithelial cells present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, of skin and endometrium, leukemia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and hemopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endometrium, melanocytes, bone marrow, adipocytes, hematopoietic cells, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and sequence similarity with tyrosine phosphatases indicate that polynucleotides and polypeptides corresponding to this gene are useful for study and treatment of cancer and hematopoietic disorders.

## **20. FEATURES OF PROTEIN ENCODED BY GENE NO: 70**

This gene is expressed primarily in osteoclastoma, breast, and infant brain and to a lesser extent in various fetal and transformed bone, ovarian, and neuronal cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: degenerative conditions of the brain and skeleton. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and skeletal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone, mammary tissue, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study and treatment of degenerative, neurological and skeletal disorders.

## 5    **FEATURES OF PROTEIN ENCODED BY GENE NO: 71**

This gene was originally cloned from tumor cell lines. Recently another group has also cloned this gene, calling it the human malignant melanoma metastasis-suppressor (KiSS-1) gene. (See Accession No. U43527.) Preferred polypeptide fragments comprise the amino acid sequence: LEKVASVGNSRPTGQQLESLGLLA (SEQ ID NO: 632); VHREEASCYCQAEPGDL (SEQ ID NO: 633); RPALRQAGGGTREPRQKRWAGL (SEQ ID NO: 634); and AVNFRPQRSQSM (SEQ ID NO: 635). Any frame shifts can easily be resolved using known molecular biology techniques.

This gene is expressed primarily in many types of carcinomas and to a lesser extent in many normal organs.

15       Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissues(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer particularly melanomas, and other hyperproliferative disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in  
20       providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of transformed organ tissue, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or  
25       another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. As a tumor suppressor gene, increase amounts of the polypeptide can be used to treat patients having a particular cancer.

30       The tissue distribution indicates that this gene and the translated product is useful for diagnosing and study of cancer.

## **FEATURES OF PROTEIN ENCODED BY GENE NO: 72**

This gene is expressed primarily in striatum and to a lesser extent in adipocytes and hemangiopericytoma.

35       Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of striatal cells present in a biological sample and for diagnosis of diseases and conditions: neurological, fat and lysosomal storage

diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., striatal tissue, adipocytes, and vascular tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis, study and treatment of neurodegenerative and growth disorders.

#### 15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 73**

This gene is expressed primarily in bone marrow stromal cells and to a lesser extent in smooth muscle, testes, endothelium, and brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of bone marrow present in a biological sample and for diagnosis of diseases and conditions: connective tissue and hematopoietic diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone marrow, stromal cells, smooth muscle, testes and other reproductive tissue, endothelium, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis, and treatment of connective tissue and blood diseases.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 74**

This gene is expressed primarily in brain, fetal liver and lung and to a lesser extent in retina, spinal chord, activated T-cells and endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of brain and regenerating liver present in a biological sample and for diagnosis of diseases and conditions: CNS and spinal chord injuries, immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, liver, pulmonary tissue, blood cells, and endothelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study and treatment of hematopoietic and neurological conditions.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 75**

The translation product of this gene shares sequence homology with GTP binding proteins (intracellular).

This gene is expressed primarily in bone marrow, brain, and melanocytes and to a lesser extent in various endocrine and hematopoietic tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hematopoietic and nervous system conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone marrow, melanocytes, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder,

relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to nucleotide binding factors indicate that polynucleotides and polypeptides corresponding to this gene are useful for study,  
5 diagnosis, and treatment of brain degenerative, skin and blood diseases.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 76**

This gene is expressed primarily in activated T-cells and to a lesser extent in retina, brain, and fetal bone.

10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of activated T-cells and developing brain present in a biological sample and for diagnosis of diseases and conditions: immune deficiencies and skeletal and neuronal growth disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes  
15 for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous, immune, and skeletomuscular systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, brain and other tissue of the nervous system, retinal tissue, and bone, and cancerous and wounded tissues) or  
20 bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides  
25 corresponding to this gene are useful for diagnosis, study and treatment of cancer, urogenital, and brain degenerative diseases.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 77**

This gene is expressed primarily in fetal liver, activated monocytes, osteoblasts  
30 and to a lesser extent in synovial, brain, and lymphoid tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of myeloid and lymphoid present in a biological sample and for diagnosis of diseases and conditions: inflammation, immune deficiencies, cancer. Similarly, polypeptides and antibodies directed to these  
35 polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and skeleton, expression of this gene at significantly



higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, blood cells, bone, synovial tissue, brain and other tissue of the nervous system, and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample  
5 taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis, and treatment of lymphoid  
10 and mesenchymal cancers and nervous system diseases.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 78**

The translation product of this gene shares sequence homology with polymerase polypeptide precursor which is thought to be important in DNA repair and replication  
15 This gene is expressed primarily in infant brain and to a lesser extent in tumors and tumor cell lines

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are  
20 not limited to, especially of the neural system and developing organs. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neural system expression of this gene at significantly higher or lower levels may be routinely detected  
25 in certain (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30 The tissue distribution and homology to polymerase polypeptide precursor indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cancers especially of the neural system and developing organs

#### **35 FEATURES OF PROTEIN ENCODED BY GENE NO: 79**

This gene is expressed primarily in muscle and endothelial cells and to a lesser extent in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: vascular diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain (e.g., muscle, endothelial cells, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of disorders of the vascular and neural system including cardiovascular and endothelial.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 80**

This gene is expressed primarily in placenta and to a lesser extent in fetal liver. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: developmental disorders and disorder of the haemopoietic system, fetal liver and placenta. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of developmental disorders and disorder of the haemopoietic system, fetal liver and placenta, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta and liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of developmental disorders and disorders of the haemopoietic system, fetal liver and placenta.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 81**

This gene is expressed primarily in bone marrow, placenta and tissues and organs of the hematopoietic system.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: disorders of the bone and haemopoietic system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification  
10 of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, bone and hematopoietic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone marrow, placenta, and hematopoietic cells, and cancerous and  
15 wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of disorders of the  
20 immune, bone and hematopoietic system

**FEATURES OF PROTEIN ENCODED BY GENE NO: 82**

The translation product of this gene shares sequence homology with secretory carrier membrane protein which is thought to be important in protein transport and  
25 export. Any frame shifts in coding sequence can be easily resolved using standard molecular biology techniques. Another group recently cloned this gene, calling it SCAMP. (See Accession No. 2232243.)

This gene is expressed primarily in prostate, breast and spleen, and to a lesser extent in several other tissues and organs.

30 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: disorders of the breast prostate and spleen. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification  
35 of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly disorders of the breast prostate and spleen, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell

types (e.g., prostate, mammary tissue, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to secretory carrier membrane protein indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of disorders of the breast, prostate and spleen.

#### 10 **FEATURES OF PROTEIN ENCODED BY GENE NO: 83**

This gene is expressed primarily in developing organs and tissue like placenta and infant brain and to a lesser extent in developed organs and tissue like cerebellum and heart.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurological diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neural system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, heart, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of diseases of the neural system including neurological disorders and cancer.

30

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 84**

The translation product of this gene shares sequence homology with ATPase 6 in *Trypanosoma brucei* which is thought to be important in metabolism.

This gene is expressed primarily in tumor and fetal tissues and to a lesser extent in melanocytes, kidney cortex, monocytes and ovary.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions: metabolism disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., fetal tissues, melanocytes, kidney, blood cells, ovary and other tissue of the reproductive system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to ATPase indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of metabolism disorders, especially in fetal and tumor tissue growth.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 85**

The translation product of this gene shares sequence homology with the immunoglobulin superfamily of proteins which are known to be important in immune response and immunity.

This gene is expressed primarily in stromal cells, colon cancer, lung, amygdala, melanocyte and to a lesser extent in a variety of other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of stromal cell development and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the stromal cells, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., stromal cells, colon, lung, amygdala, and melanocytes, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to immunoglobulin indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of immune system disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 86**

The translation product of this gene shares sequence homology with transcription initiation factor eIF-4 gamma which is thought to be important in gene transcription.

This gene is expressed primarily in tumor tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumorigenesis.

Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly in tumor tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endometrium and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to transcription initiation factor eIF-4 gamma indicate that polynucleotides and polypeptides corresponding to this gene are useful for gene regulation in tumorigenesis.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 87**

The translation product of this gene shares sequence homology at low level in prolines with secreted basic proline-rich peptide II-2 which is thought to be important in protein structure or inhibiting hydroxyapatite formation in vitro.

This gene is expressed primarily in endometrial tumor and fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: endometrial tumors.

Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the muscular/skeletal and reproductive systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endometrium, and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample

taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- 5 The tissue distribution and homology to secreted basic proline-rich peptide II-2 indicate that polynucleotides and polypeptides corresponding to this gene are useful for inhibiting hydroxyapatite formation or establishing cell/tissue structure.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 88**

- 10 This gene is expressed primarily in: amniotic cells induced with TNF in culture; and to a lesser extent in colon tissue from a patient with Crohn's Disease; parathyroid tumor; activated T-cells; cells of the human Caco-2 cell line; adenocarcinoma; colon; corpus colosum; fetal kidney; pancreas tumor; fetal brain; early stage brain, and anergic T-cells.

- 15 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system; 20 e.g., tumors, expression of this gene at significantly higher or lower levels may be routinely detected in certain (e.g., amniotic cells, colon, kidney, pancreas, parathyroid, brain and other tissue of the nervous system, blood cells, hematopoietic cells, liver, spleen, bone, testes and other reproductive tissue, brain and other tissue of the nervous system, and epithelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., 25 serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- 30 The tissue distribution indicates that the protein product of this clone is useful for modulating tumorigenesis and other immune system conditions such as disorders in immune response.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 89**

- 35 This gene is expressed primarily in fetal liver/spleen and hematopoietic cells and to a lesser extent in brain, osteosarcoma, and testis tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions: leukemia and hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., hematopoietic cells, liver, spleen, bone, testes, and other reproductive tissue, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of hematopoietic and immune disorders.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 90**

The translation product of this gene shares weak sequence homology with mouse Gcap1 protein which is developmentally regulated in brain.

This gene is expressed primarily in infant and adult brain and fetal liver/spleen and to a lesser extent in smooth muscle, T cells, and a variety of other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurological or hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous, hematopoietic, immune, and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, blood cells, liver, spleen, and smooth muscle, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.



The tissue distribution and its homology to Gcap1 protein indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of disorders in neuronal, hematopoietic, immune, and endocrine systems.

## 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 91

This gene is expressed primarily in brain and hematopoietic cells and to a lesser extent in tumor tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: disorder in nervous, hematopoietic, immune systems and tumorigenesis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous, hematopoietic, immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for diagnosis and treatment of disorders in the nervous, hematopoietic, and immune systems.

25

## FEATURES OF PROTEIN ENCODED BY GENE NO: 92

The translation product of this gene shares sequence homology with neuroendocrine-specific protein A which is thought to be important in neurologic systems.

30 This gene is expressed primarily in brain tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neural disorders and degeneration disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central or peripheral nervous systems, expression of this gene at

35

significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., hematopoietic cells, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to neuroendocrine-specific protein A indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment or diagnosis of neural disorders and degeneration disease.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 93**

The translation product of this gene shares sequence homology with collagen-like protein and prolin-rich protein which are thought to be important in connective tissue function and tissue structure.

This gene is expressed primarily in fetal liver/spleen and brain tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neuronal or hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, spleen, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to collagen-like protein and proline-rich proteins indicate that polynucleotides and polypeptides corresponding to this gene are useful for supporting brain and hematopoietic tissue function and diagnosis and treatment of disorders in these functions.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 94**

This gene is expressed primarily in embryonic tissues and tumor tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions which include, but are not limited to,. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system (e.g., tumors), expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic tissue and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cancer.

#### 15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 95**

This gene is expressed primarily in brain tumor, placenta, and melanoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: brain tumor or melanoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain or melanocytes, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, placenta, and melanocytes, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the translation product of this gene is useful in the diagnosis and treatment of brain tumors and melanoma.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 96**

The translation product of this gene shares sequence homology with a yeast membrane protein, SUR4, which encodes for APA1 that acts on a glucose-signaling pathway that controls the expression of several genes that are transcriptionally regulated by glucose.

This gene is expressed primarily in fetal liver, and to a lesser extent in placenta and breast tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of fetal liver or defects of glucose-regulated ATPase activities in tissues. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetal immune/hematopoietic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, placenta, and mammary tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to yeast SUR4 membrane protein indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of defects of fetal liver or defects of glucose-regulated ATPase activities.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 97**

This gene is expressed primarily in fetal liver, brain, and amniotic fluid.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of the fetal immune system and adult brain. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetal immune system and adult brain, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., amniotic fluid, serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for detecting defects of the fetal immune and hematopoietic systems since fetal liver is

the predominant organ responsible for hematopoiesis in the fetus. In addition, the gene product of this gene is thought to be useful for detecting certain neurological defects of the brain.

## 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 98

The translation product of this gene shares sequence homology with an yolk protein precursor, Vitellogenin which is thought to be important in binding lipids such as phosvitin.

This gene is expressed primarily in amniotic cells and fetal liver.

10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects in amniotic cells, fetal liver development and the fetal immune system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes  
15 for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the [insert system where a related disease state is likely, e.g., immune], expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., amniotic cells, and liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,  
20 urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to vitellogenin indicate that the protein  
25 product of this clone is useful for treatment and diagnosis of defects in amniotic cells, fetal liver development and the fetal immune system.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 99

This gene is expressed primarily in placenta, endometrial tumor, osteosarcoma  
30 and stromal cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumor of the endometrium or bone, and osteosarcoma. Similarly, polypeptides and antibodies  
35 directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the obstetric system (e.g. placenta,

endometrium) and the bones, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, endometrium, bone, and stromal cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of tumors and abnormalities of the endometrium, and the bones because of its abundance in the aforementioned tissues..

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 100**

This gene is expressed primarily in hepatocellular tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hepatocellular tumor. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the liver, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for diagnosis and treatment of hepatocellular cancer because of its abundant expression in this tissue.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 101**

This gene is expressed primarily in Corpus Colosum, fetal lung and infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of the Corpus Colosum or defects of the fetal lung. Similarly, polypeptides and antibodies directed to

these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the Corpus Colosum and brain in general, and fetal lung, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., lung, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for diagnosis and treatment of defects of the Corpus Colosum and brain in general, and defects of fetal lung.

#### 15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 102**

This gene is expressed primarily in T cells and stromal cells, and to a lesser extent in adrenal gland.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of T cell immunity and stromal cell development. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, stromal cells, and adrenal gland, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for diagnosis and treatment of defects of T cell immunity and stromal cell development because of its abundant expression in these tissues.

#### 35 **FEATURES OF PROTEIN ENCODED BY GENE NO: 103**

This gene is expressed primarily in infant brain and placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of the brain and nervous system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, especially brain, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and placenta, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for detecting defects of the brain, especially in young children.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 105**

This gene is expressed primarily in human osteoclastoma and to a lesser extent in human pancreas tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer particularly osteoclastoma and pancreatic tumor. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly in transformed tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone and pancreas, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for diagnosis and treatment of some types of tumors, particularly pancreatic cancer and osteoclastoma.



**FEATURES OF PROTEIN ENCODED BY GENE NO: 106**

This gene is expressed primarily in fetal liver/spleen, and to a lesser extent in activated T-Cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, spleen, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis or treatment of immune disorders.

**20 FEATURES OF PROTEIN ENCODED BY GENE NO: 107**

This gene is expressed primarily in human embryo and to a lesser extent in spleen and chronic lymphocytic leukemia.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: leukemia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune or hemopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic tissue, spleen, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for the diagnosis and treatment of leukemia.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 108**

This gene is expressed primarily in placenta, and to a lesser extent in early stage human brain and in lung.

- 5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: fetal developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly in fetal and amniotic tissue, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, brain and other tissue of the nervous system, and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another
- 10 tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- The tissue distribution indicates that the protein product of this is useful for production of growth factor(s) associated with fetal development. Preferred
- 20 polypeptides comprise the full-length polypeptide shown in the sequence listing, truncated however, at the amino terminus and beginning with QTIE.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 109**

- This gene is expressed primarily in fetal spleen, and to a lesser extent in B-Cell lymphoma and T-Cell lymphoma.
- 25

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing
- 30 immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., spleen and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.
- 35

The tissue distribution indicates that the protein product of this clone is useful for the treatment and diagnosis of human lymphomas.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 110**

5           The translation product of this gene shares sequence homology with sarcoma amplified sequence (SAS), a tetraspan receptor which is thought to be important in malignant fibrous histiocyoma and liposarcoma.

This gene is expressed primarily in human osteoclastoma, and to a lesser extent in pineal gland and infant brain.

10           Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: malignant fibrous histiocyoma and liposarcoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification  
15 of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone, pineal gland, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal  
20 fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to sarcoma amplified sequence (SAS) indicate that the protein product of this clone is useful for treatment of, osteosarcoma,  
25 malignant fibrous histiocyoma and liposarcoma and related cancers, particularly sarcomas.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 111**

30           The translation product of this gene shares sequence homology with 6.8K proteolipid protein, mitochondrial - bovine.

This gene is expressed primarily in Wilm's tumor and to a lesser extent in cerebellum and placenta.

35           Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: Wilm's tumor. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell

type(s). For a number of disorders of the above tissues or cells, particularly of the immune or renal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to 6.8K proteolipid protein indicate that the protein product of this clone is useful for diagnostic and therapeutics associated with tumors, particularly Wilm's tumor disease.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 112**

This gene is expressed primarily in embryonic tissue and to a lesser extent in osteoblasts, endothelial cells, macrophages (GM-CSF treated), and bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic tissue, bone, endothelial cells, blood cells and bone marrow, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment or diagnosis of immune disorders. Preferred polypeptides encoded by this gene comprise the following amino acid sequence: MITDVQLAIFANMLGVSLFLLVVLYHYVAVNNPKKQE (SEQ ID NO: 636).

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 113**

This gene is expressed primarily in hepatocellular tumor, and to a lesser extent in fetal liver/spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors, particularly hepatocellular tumors. Similarly, polypeptides and antibodies directed to these

5 polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hepatic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,

10 urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful

15 for diagnosis and treatment of tumors, particularly hepatocellular tumors.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 114**

The translation product of this gene exhibits a very high degree of sequence identity with the human Pig8 gene which is thought to be important in p53 mediated

20 apoptosis. The sequence of this gene has since been published by Polyak and colleagues (Nature 389, 300-306 (1997)). In addition, the predicted translation product of this contig exhibits very high sequence homology with a murine gene denoted as EI24 which is also thought to be important in p53 mediated apoptosis.

This gene is expressed primarily in infant brain and activated T-cells and to a

25 lesser extent in bone marrow, fetal liver, and prostate.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, and tissue damage by radiation and anti-cancer drugs. Similarly,

30 polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the

35 nervous system, blood cells, bone marrow, liver, and prostate, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder,

relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to human Pig8 and murine EI24 genes indicate that polynucleotides and polypeptides corresponding to this gene are useful for preventing apoptosis in patients being treated with anti-oncogenic drugs such as etoposide, hydroperoxycyclophosphamide, and X-irradiation, since this protein product is upregulated in cells undergoing such treatment where p53 was overexpressed. It may also be useful in the treatment of hematopoietic disorders and in boosting numbers of hematopoietic stem cells by interfering with the apoptosis of progenitor cells. The mature polypeptide is predicted to comprise the following amino acid sequence:

EEMADSVKTLQDLARGIKDSIWGICTISKLDARIQQKREEQRRRRASSVLAQRRRAQSIERKQES  
 EPRIVSRIFQCCA WNGGVFWFSLLLFYRVFIPVLQSVTARIIGDPSLHGDVWSWLEFFLTSIFSA  
 LWVLPFLVLSKVVNAIWFQDIADLAFEVSGRKPFPFSVSKIIADMLFNLLLQALFLIQGMFVSL  
 FPIHLVGQLVSLHMSLLYSLYCFEYRWFNKGIEHQRLSNIERNWPYYFGFGLPLAFLTAMQ  
 SSYIISGCLFSILFPLFIISANEAKTPGKAYLFQLRLFSLVVFLSNRLFHKTVYLQSALSSSTS  
 AEFSPHPSPAKLKATAGH (SEQ ID NO: 637). Accordingly, polypeptides comprising the foregoing amino acid sequence are provided as are polynucleotides encoded such polypeptides.

## 20 FEATURES OF PROTEIN ENCODED BY GENE NO: 115

This gene is expressed primarily in stromal cells and to a lesser extent in multiple sclerosis.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: affecting the nervous system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., stromal cells and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of multiple sclerosis and other autoimmune diseases.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 116**

This gene is expressed primarily in the gall bladder

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: gall stones or infection of the digestive system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system or renal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., gall bladder and tissue of the digestive system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for possible prevention of digestive disorders where there may be a lack of digestive enzymes produced or in the detection and possible prevention of gall stones.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 117**

The translation product of this gene shares sequence homology with dystrophin gene which is thought to be important in building and maintenance of muscles.

This gene is expressed primarily in placenta and to a lesser extent in fetal brain and fetal liver, and spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: muscular dystrophy, Duchenne and Becker's muscular dystrophies. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal muscle system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, brain and other tissue of the nervous system, muscle, liver, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from

an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- 5 The tissue distribution and homology to the dystrophin gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diseases related the degenerative myopathies that are characterized by the weakness and atrophy of muscles without neural degradation; such as Duchenne and Becker's muscular dystrophies.

#### 10 **FEATURES OF PROTEIN ENCODED BY GENE NO: 118**

This gene is expressed primarily in olfactory tissue and to a lesser extent in cartilage.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: connective tissue diseases; chondrosarcoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the connective tissue, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., olfactory tissue and cartilage, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for tumors of connective tissues, osteoarthritis and the treatment and diagnosis of chondrosarcoma.

#### 30 **FEATURES OF PROTEIN ENCODED BY GENE NO: 119**

This gene is expressed primarily in Activated Neutrophils and to a lesser extent in fetal spleen, and CD34 positive cells from cord blood.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: allergies, defects in hematopoiesis and inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential



identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and hematopoiesis system the, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and spleen, and cancerous and  
5 wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides  
10 corresponding to this gene are useful for reducing the allergic effects felt by allergy suffers by neutralizing the activity of the immune system, especially since neutrophils are abundant in persons suffering from allergies and other inflammatory conditions.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 120**

15 The translation product of this gene shares sequence homology with poly A binding protein II which is thought to be important in RNA binding for transcription of RNA to DNA

This gene is expressed primarily in colon and to a lesser extent in brain and immune system.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: colon cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a  
25 number of disorders of the above tissues or cells, particularly of the immune and digestive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., colon, tissue and cells of the immune system, and brain or other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal  
30 fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to poly A binding protein II indicate that polynucleotides and polypeptides corresponding to this gene are useful for detection  
35 and treatment of colon cancer and other disorders of the digestive system..

**FEATURES OF PROTEIN ENCODED BY GENE NO: 121**

The translation product of this gene shares sequence homology with thymidine diphosphoglucose 4.6 dehydrase which is thought to be important in the metabolism of sugar.

- 5        This gene is expressed primarily in fetal liver and spleen and to a lesser extent in infant brain.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diabetes. Similarly,
- 10       polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, spleen, and brain and other tissue of the
- 15       nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- 20       The tissue distribution and homology to thymidine diphosphoglucose 4.6 dehydrase indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment of persons with diabetes since it appears that this protein is needed in the metabolism of sugar in to its more basic components.

25       **FEATURES OF PROTEIN ENCODED BY GENE NO: 122**

- The translation product of this gene shares sequence homology with ceruloplasmin which is thought to be important in the metabolism and transport of iron and copper. Ceruloplasmin also contains domains with homology to clotting factors V and VIII. Defects in the circulating levels of ceruloplasmin (aceruloplasminemia) have
- 30       been associated with certain disease conditions such as Wilson disease, and the accompanying hepatolenticular degeneration.

         This gene is expressed primarily in brain and retina and to a lesser extent in endothelial cells.

- Therefore, polynucleotides and polypeptides of the invention are useful as
- 35       reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diseases marked by defects in iron metabolism; aceruloplasminemia not characterized by defects in the

known ceruloplasmin gene locus; nonclassical Wilson disease; movement disorders; and tumors derived from a brain tissue origin. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, retina, and nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, retinal tissue, and endothelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to ceruloplasmin indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment of patients with aceruloplasminemia, or other defects in iron and/or copper metabolism. Mutations in this locus could also be diagnostic for patients currently experiencing or predicted to experience aceruloplasminemia.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 123**

This gene is expressed primarily in brain and B cell lymphoma and to a lesser extent in fetal liver and spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: B cell lymphoma; tumors and diseases of the brain and/or spleen; hematopoietic defects. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and hematopoietic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, blood cells, liver, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of disorders in neuronal,

hematopoietic, and immune systems. It could potentially be useful for neurodegenerative disorders and neuronal and/or hematopoietic cell survival or proliferation.

#### 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 124

This gene is expressed primarily in osteoclastoma, dermatofibrosarcoma, and B cell lymphoma and to a lesser extent in endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer in particular osteoclastoma, dermatofibrosarcoma, and B cell lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the bone, immune, and circulatory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone, epidermis, blood cells, and endothelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cancers and lymphoma; osteoporosis; and the control of cell proliferation and/or differentiation.

#### 25 FEATURES OF PROTEIN ENCODED BY GENE NO: 125

This gene is expressed primarily in immune tissues and hematopoietic cells, particularly in activated T cells and neutrophils, spleen, and fetal liver, and to a lesser extent in infant adrenal gland.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects in T cell activation; hematopoietic disorders; tumors of a hematopoietic and/or adrenal gland origin. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic and/or endocrine systems, expression of this gene at significantly higher

- or lower levels may be routinely detected in certain tissues and cell types (e.g., cells and tissues of the immune system, hematopoietic cells, blood cells, liver, and adrenal gland, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual
- 5 having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for immune and/or hematopoietic disorders;
- 10 diseases related to proliferation and/or differentiation of hematopoietic cells; defects in T cell and neutrophil activation and responsiveness; and endocrine and/or metabolic disorders, particularly of early childhood.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 126**

- 15 This gene is expressed primarily in placenta and endothelial cells and to a lesser extent in melanocytes and embryonic tissues.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors of an endothelial
- 20 cell origin; angiogenesis associated with tumor development and metastasis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vascular system and developing embryo, expression of this gene at significantly higher or lower levels
- 25 may be routinely detected in certain tissues and cell types (e.g., placenta, endothelial cells, melanocytes, and embryonic tissues, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily
- 30 fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of developmental disorders; inhibition of angiogenesis; and vascular patterning.

#### **35 FEATURES OF PROTEIN ENCODED BY GENE NO: 127**

This gene is expressed primarily in endothelial cells and hematopoietic tissues, including spleen, tonsils, leukocytes, and both B- and T-cell lymphomas.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors of an endothelial cell and/or hematopoietic origin; leukemias and lymphomas. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and vascular systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endothelial cells, hematopoietic cells, spleen, tonsils, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the manipulation of angiogenesis; the differentiation and morphogenesis of endothelial cells; the proliferation and/or differentiation of hematopoietic cells; and the commitment of hematopoietic cells to distinct cell lineages.

20

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 128**

This gene is expressed primarily in kidney medulla and to a lesser extent in spleen from chronic myelogenous leukemia patients, prostate cancer, and some other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors of a kidney origin; chronic myelogenous leukemia; prostate cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the kidney and spleen, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., kidney, spleen, and prostate, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of kidney disorders and cancer, particularly chronic myelogenous leukemia and prostate cancer. It may also be useful for the enhancement of kidney tubule regeneration in the treatment of acute renal failure.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 129**

This gene is expressed primarily in adult and infant brain and to a lesser extent in mesenchymal or fibroblast cells, as well as tissues with a mesenchymal origin.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors of a brain and/or mesenchymal origin; neurodegenerative disorders; cancer; fibrosis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and of mesenchymal cells and tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis of tumors of a brain and/or mesenchymal origin; neurodegenerative disorders; cancer; and fibrosis, based upon the expression of this gene within those tissues. Fibrosis is considered as mesenchymal cells and fibroblasts are the primary cellular targets involved in this pathological condition.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 130**

This gene is expressed primarily in hepatocellular cancer and to a lesser extent in fetal tissues as well as testes tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: liver cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing

immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, fetal tissue, and testes and other  
5 reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10       The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of liver cancer.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 131**

      This gene is expressed only in infant early brain.

15       Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: development and diseases of the nervous system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification  
20 of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another  
25 tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

      The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating diseases of the brain in children and in  
30 treating nervous system disorders such as Alzheimer's disease, schizophrenia, dementia, depression, etc.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 132**

      This gene is expressed primarily in brain and to a lesser extent in glioblastoma.

35       Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: Alzheimer's disease,



schizophrenia, depression, mania, and dementia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating brain disorders such as Alzheimer's disease, schizophrenia, depression, mania, and dementia.

#### 15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 133**

The translation product of this gene shares sequence homology with ribitol dehydrogenase of bacteria which is thought to be important in metabolism of sugars.

This gene is expressed primarily in macrophage and to a lesser extent in T-cell lymphoma and lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tissue destruction in inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to ribitol dehydrogenase indicate that polynucleotides and polypeptides corresponding to this gene are useful for altering macrophage metabolism in diseases such as inflammation where macrophages are causing excess tissue destruction.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 134**

This gene is expressed primarily in pancreatic tumor and to a lesser extent in synovial sarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to,. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine and connective tissue systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., pancreas, and synovial tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating and diagnosing various cancers.

**20 FEATURES OF PROTEIN ENCODED BY GENE NO: 135**

This gene is expressed primarily in T cell lines such as Raji and to a lesser extent in infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune system disorders and inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating and diagnosing inflammatory diseases

such as rheumatoid arthritis, sepsis, inflammatory bowel disease, and psoriasis, as well as neutropenia.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 136**

5       The translation product of this gene shares high sequence homology with SAR1 subfamily of GTP-binding proteins which is thought to be important in vesicular transport in mammalian cells.

      This gene is expressed primarily in serum-stimulated smooth muscle cells and to a lesser extent in a T-cell lymphoma.

10       Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diseases affecting vesicular transport. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification  
15 of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the muscular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and smooth muscle, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample  
20 taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

      The tissue distribution and homology to GTP-binding proteins indicate that polynucleotides and polypeptides corresponding to this gene are useful for gene therapy  
25 in treating the large number of diseases involved in defective vesicular transport within cells..

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 137**

      The translation product of this gene shares sequence homology with a protein  
30 found in *C. elegans* cosmid F25B5.

      This gene is expressed primarily in a fetal tissues and to a lesser extent in melanocytes.

      Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a  
35 biological sample and for diagnosis of diseases and conditions: abnormal fetal development, especially of the pulmonary system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes

for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetal pulmonary system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., fetal tissue, pulmonary tissue, and melanocytes, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of diseases affecting the pulmonary system, such as emphysema.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 138**

This gene is expressed primarily in gall bladder and to a lesser extent in smooth muscle.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: digestive system disease and gall bladder problems. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., gall bladder and tissue of the digestive system, and smooth muscle, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating diseases of the digestive system.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 139**

This gene is expressed primarily in placenta and to a lesser extent in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: abnormal fetal development. Similarly, polypeptides and antibodies directed to these polypeptides are

useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of developing tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, and brain and other  
5 tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- 10       The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating and diagnosing abnormal fetal development.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 140**

- 15       This gene is expressed primarily in smooth muscle and to a lesser extent in ovary, prostate cancer, and activated monocytes.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hypertension and  
20 atherosclerosis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the circulatory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., smooth  
25 muscle, ovary and other reproductive tissue, prostate, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- 30       The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating diseases of the circulatory system, such as hypertension, atherosclerosis, etc.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 141**

- 35       This gene is expressed primarily in fetal spleen and to a lesser extent in placenta and bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: anemia and other diseases affecting blood cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the circulatory and pulmonary systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., spleen, placenta, bone marrow, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the generation of red and white blood cells and for the diagnosis of disease of these cells.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 142**

The predicted translation product of this contig is a human homolog of the murine tetracycline/sugar transporter molecule recently reported by Matsuo and colleagues (Biochem. Biophys. Res. Commun. 238 (1), 126-129 (1997)).

This gene is expressed primarily in synovium and to a lesser extent in endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: rheumatoid arthritis and inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and lymphatic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., synovial tissue, and endothelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of inflammatory diseases, such as rheumatoid arthritis, leukemia, neutropenia, inflammatory bowel disease, psoriasis, sepsis, and the like.

5

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 143**

This gene is expressed primarily in placenta and to a lesser extent in melanocyte, fetal liver and spleen, and bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as  
10 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: abnormal early development. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, lower levels  
15 may be routinely detected in certain tissues and cell types (e.g., placenta, melanocytes, liver, spleen, and bone marrow, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an  
20 individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of abnormal early development phenomena and diseases.

#### **25 FEATURES OF PROTEIN ENCODED BY GENE NO: 144**

This gene is expressed primarily in fetal liver and spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: anemia and neutropenia.  
30 Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and blood systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver and spleen,  
35 and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the

expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful in hematopoiesis and bone marrow regeneration as it is most abundant in fetal tissues responsible for the generation of hematopoietic cells.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 145**

The translation product of this gene shares sequence homology with protein tyrosine phosphatase which is thought to be important in transducing signal to activate cells such as T cell, B cell and other cell types.

This gene is expressed primarily in T cells and tissues in early stages of development and to a lesser extent in cancers.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immuno-related diseases and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic and fetal tissue, undifferentiated cells, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to the protein tyrosine phosphatase family indicate that polynucleotides and polypeptides corresponding to this gene are useful for modulating the immune system.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 146**

This gene is expressed primarily in T cell and to a lesser extent in B cell, macrophages and tumor tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immuno-disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in



providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for regulating the immune system therefore can be used in treating diseases such as autoimmune diseases and cancers.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 147**

This gene is expressed primarily in placenta and to a lesser extent in endothelial cells, testis tumor, ovarian cancer, uterine cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, endothelial cells, testis and ovary and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cancers.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 148**

This sequence has significant homology to mouse torsin A. Recently, another group cloned the human Torsin A gene. (See, Accession No. 2358279; see also Nature Genet. 17, 40-48 (1997).)

This gene is expressed primarily in osteoclastoma, T-cell, and placenta and to a lesser extent in fetal lung, fetal liver, fetal brain, adult brain and tumor tissues

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: disease conditions in hematopoiesis and cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoiesis system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, bone, placenta, lung, liver, and brain and other tissues of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating blood related diseases such as deficiencies in red blood cell, white blood cell, platelet and other hematopoiesis cells.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 149**

This gene is expressed primarily in T cell, prostate and prostate cancer, endothelial cells and to a lesser extent in monocyte, dendritic cell, bone marrow, salivary gland, colon cancer, stomach cancer, pancreatic tumor, uterine cancer, fetal spleen and osteoclastoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immuno-related diseases and cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, prostate, endothelial cells, dendritic cells, bone marrow, salivary gland, colon, stomach, pancreas, uterus, spleen and bone, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of cancers.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 150

- 5 This gene was recently cloned by another group, calling it eIF3-p66. (See Accession No. 2351378.) This gene plays a role in RNA binding and macromolecular assembly, and therefore, any mutations in this gene would likely result in a diseased phenotype. Preferred polypeptide fragments comprise the amino acid sequence:
- 10 MAKFMTPIQDNPSGWGPCAVPEQFRDMPYQPFSGDRLGKVADWTGATYQDKRYTNKYSS  
QFGGGSQYAYFHEEDESSFLVDTARTQKTAYQRNRMFAQRNLRRDKDRRNMLQFNLQILP  
KSAKQKERERIRLQKKFQKQFGVRQKWDQKSQKPRDSSVEVRSDWEVKEEMDFPQLMKMRY  
LEVSEPDIECCGALEYDKAFDRIITRSEKPLRXXKRIFHTVTTTDDPVIRKLAKTQGNVFATD  
AILATLMSCTRSVYSWDIVVQRVGSKLFFDKRDNDFDLTVSETANEPPQDEGNSFNSPRNL  
AMEATYINHNFSQQCLRMGKERYNFPNPNPFVEDDMDKNEIASVAYRYRSGKLGDDIDLIVRC  
15 EHDGVMTGANGEVSFINIKTLNEWDSRHCNGVDWRQKLDSSQRGAVIATELKNNSYKLARWTC  
CALLAGSEYKLGYSRYHVKDSSRHVILGTQQKPNEFASQINLSVENAWGILRCVIDICMKL  
EEGKYLIKDPNKQVIRVYSLPDGTFSS (SEQ ID NO: 638), as well as N-terminal and C-terminal deletions of this polypeptide fragment.

- 20 This gene is expressed primarily in T cell, bone marrow, embryo and endothelial cells and to a lesser extent in testis tumor and endometrial tumor.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune diseases and tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful
- 25 in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial
- 30 fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for immune disorders and cancers.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 151**

This gene is expressed primarily in testis and to a lesser extent in T cell, spinal cord, placenta, neutrophil and monocyte.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: male reproductive and endocrine disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive, immune and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., testis and other reproductive tissue, blood cells, tissue of the nervous system, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for regulating immune and reproductive functions.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 152**

The translation product of this gene shares sequence homology with tyrosyl-tRNA synthetase which is thought to be important in cell growth.

This gene is expressed primarily in brain, liver, keratinocytes, tonsils, and heart.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer autoimmune diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, liver, keratinocytes, tonsils, heart expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissues of the nervous system, liver, keratinocytes, tonsils and heart, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard

gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to tyrosyl-tRNA synthetase indicate that polynucleotides and polypeptides corresponding to this gene are useful for modulating cell growth.

### FEATURES OF PROTEIN ENCODED BY GENE NO: 153

This gene is homologous to the *Drosophila* transcriptional regulator dre4. (See Accession No. 2511745.) Dre4 is a gene required for steroidogenesis in *Drosophila melanogaster* and encodes a developmentally expressed homologue of the yeast transcriptional regulator CDC68. Preferred polypeptide fragments comprise the amino acid sequence: KKRHTDVQFYTEVGEITDGLGKHQMHDRDDLYAEQMEREMRHKLTAFKN FIEKVEALTKEELEFEVPRDLGFNGAPYRSTCLLQPTSSALVNATEWPPFVVTLDEVELIHFXR VQFHLKNFDMVIVYKDYSKKVTMNAIPVASLDPIKEWLNSCDLKYTEGVQSLNWTIMKTIVD DPEGFFEQGGWSFL (SEQ ID NO: 639), as well as N-terminal and C-terminal deletions of this fragments. Also preferred are polynucleotide fragments encoding this polypeptide fragment.

This gene is expressed primarily in fetal liver, spleen, placenta, lung, T cell, thyroid, testes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: brain tumor, heart and liver diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetal liver, spleen, placenta, lung, T cell, thyroid, testes expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, spleen, placenta, lung, blood cells, thyroid, and testes and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

### FEATURES OF PROTEIN ENCODED BY GENE NO: 154

This gene is expressed primarily in brain and to a lesser extent in fetal heart, testis, spleen, lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: heart, liver and spleen diseases, immunological diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, fetal heart, testis, spleen, lung expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, heart, testes and other reproductive tissue, spleen, and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 155**

Activation of T cells through the T cell antigen receptor (TCR) results in the rapid tyrosine phosphorylation of a number of cellular proteins, one of the earliest being a 100 kDa protein. This gene is the human equivalent of murine valosin containing protein (VCP). VCP is a member of a family of ATP binding, homo-oligomeric proteins, and the mammalian homolog of *Saccharomyces cerevisiae cdc48p*, a protein essential to the completion of mitosis in yeast. Both endogenous and expressed murine VCP are tyrosine phosphorylated in response to T cell activation. Thus we have identified a novel component of the TCR mediated tyrosine kinase activation pathway that may provide a link between TCR activation and cell cycle control.

This gene is expressed primarily in brain, liver, spleen, placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, liver, spleen, placenta expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, liver, spleen, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from

an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

5 The tissue distribution and homology to VCR indicate that polynucleotides and polypeptides corresponding to this gene are useful for treating cancer.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 156**

10 The translation product of this gene shares sequence homology with rat growth response protein which is thought to be important in cell growth. A group recently cloned the human homolog of this gene, calling it insulin induced protein 1. (See Accession No. 2358269, see also, Genomics 43 (3), 278-284 (1997).) Preferred polypeptide fragments comprise the amino acid sequence: RSGLGLGITIAFLATLITQF LVYNGVYQYTSPDFLYIRSWLPCIFFSGGVTVGNIGRQLAMGVPEKPHSD (SEQ ID NO: 640), as well as N-terminal and C-terminal deletions of this polypeptide fragment. Also  
15 preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in brain, liver, placenta, heart, spleen, lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a  
20 biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, liver, placenta, heart, spleen.  
25 expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, liver, placenta, heart, spleen, and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the  
30 standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to growth-response protein indicate that polynucleotides and polypeptides corresponding to this gene are useful for modulating cell growth.

35

**FEATURES OF PROTEIN ENCODED BY GENE NO: 157**

This gene is expressed primarily in Glioblastoma, endometrial tumor, lymphoma and pancreas tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: Glioblastoma, Endometrial tumor, lymphoma and pancreas tumor. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endometrium, lymphoid tissue, pancreas, and tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 158**

The translation product of this gene shares sequence homology with IGE receptor which is thought to be important in allergy and asthma.

This gene is expressed primarily in T cell, and fetal liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: allergy and asthma and other immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.



The tissue distribution and homology to IgE receptor indicate that polynucleotides and polypeptides corresponding to this gene are useful for allergy and asthma.

#### 5    **FEATURES OF PROTEIN ENCODED BY GENE NO: 159**

The translation product of this gene shares sequence homology with immunoglobulin heavy chain which is thought to be important in immune response to the antigen.

10    This gene is expressed primarily in activated neutrophil and to a lesser extent in activated T cell, monocyte and heart.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: infection, inflammation and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are  
15    useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and heart, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial  
20    fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to immunoglobulin heavy chain variable region indicate that polynucleotides and polypeptides corresponding to this gene are  
25    useful for making the ligand to block specific antigen which cause certain disease.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 160**

The translation product of this gene shares sequence homology with mouse X inactive specific transcript protein which is thought to be important in X chromosome  
30    inactivation.

This gene is expressed primarily in HSA172 cell and to a lesser extent in normal ovary tissue, ovarian cancer, frontal cortex and brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a  
35    biological sample and for diagnosis of diseases and conditions: ovarian tumor, schizophrenia and other neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for

differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and neural system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., ovary and other reproductive tissue, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to X inactive specific transcript protein indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of reproductive system tumors and CNS tumors.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 161**

This gene is expressed primarily in adipose cell and to a lesser extent in liver and prostate.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: obesity and liver disorder. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the adipose cell, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., adipose cells, liver, and prostate, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of obesity and liver disorder.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 162**

The translation product of this gene shares sequence homology with yeast ubiquitin activating enzyme homolog which is thought to be important in protein posttraslation processing.

This gene is expressed primarily in stromal cell and to a lesser extent in retina, H. Atrophic Endometrium, colon carcinoma and myeloid progenitor cell.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of stromal cell development, neuronal growth disorders and tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and neural system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., retinal cells, endometrium, colon, and bone marrow, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to ubiquitin-activating enzyme homolog indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis or treatment of some type of tumors, fucosidosis and neuronal growth disorders.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 163**

This gene is expressed primarily in primary breast cancer and hemangiopericytoma and to a lesser extent in adult brain and cerebellum.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: breast cancer, leukemia and cerebellum disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and neural system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., mammary tissue, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis or treatment of various tumors and disease involved in neural system.

**5 FEATURES OF PROTEIN ENCODED BY GENE NO: 164**

The translation product of this gene shares sequence homology with proline rich proteins. Recently, another group has also cloned this gene, calling it CD84 leukocyte antigen, a new member of the Ig superfamily. (See Accession No. U82988, see also, Blood 90 (6), 2398-2405 (1997).)

10 This gene is expressed primarily in Weizmann olfactory tissue and osteoclastoma and to a lesser extent in anergic T-cell.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: osteoporosis and immune  
15 disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., olfactory tissue, bone, and  
20 blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25 The tissue distribution and homology to the Ig superfamily indicate that the protein product of this clone is useful for treatment of osteoporosis, autoimmune disease, and other immune disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 165**

30 This gene is expressed primarily in atrophic endometrium and colon cancer and to a lesser extent in some fetal tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors. Similarly,  
35 polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system,

expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endometrium, colon, and fetal tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of tumors, specifically endometrium and colon tumors.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 166**

This gene is expressed primarily in human primary breast cancer and to a lesser extent in activated monocyte. Although the predicted signal sequence is identified in Table 1, other upstream sequences are also relevant. Preferred polypeptide fragments comprise the amino acid sequence: VTQPKHLSASMGGSEIPFSFYYPWELAXXPXVRISWRRGHFHG QSFYSTRPPSIHKDYVNRLFLNWTEGQESGFLRISNLRKEDQSVYFCRVELDTRRSG (SEQ ID NO: 641), as well as N-terminal and C-terminal deletions. Also preferred are polynucleotide fragments encoding these polypeptide fragments.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: breast cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., mammary tissue, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis of breast cancer.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 167**

This gene is expressed primarily in fetal tissues and to a lesser extent in adult lung. This gene has also been mapped to chromosomal location 9q34, and thus, can be used as a marker for linkage analysis for chromosome 9.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the embryo tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., fetal tissues, and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 168**

The translation product of this gene shares sequence homology with Ig Heavy Chain which is thought to be important in immune response.

This gene is expressed primarily in prostate cancer tissue specifically

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: prostate cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the prostate, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., prostate, tissue and cells of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 169**

The translation product of this gene shares sequence homology with cytosolic acyl coenzyme-A hydrolase, which is thought to be important in neuron-specific fatty acid metabolism. The gene represented by this contig has since been published by Hajra and colleagues (GenBank Accession No. U91316).

This gene is expressed primarily in human pituitary gland and to a lesser extent in colorectal cancer tissue. This gene has also been observed in the LNCAP cell line.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hyperlipidemias of familial and/or idiopathic origins. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly blood, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., pituitary and colon, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to rat cytosolic acyl coenzyme-A hydrolase indicate that polynucleotides and polypeptides corresponding to this gene are useful for the detection or treatment of hyperlipidemia disease states by virtue of the ability of specific drugs to activate the enzyme.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 170

The translation product of this gene shares sequence homology with a *Caenorhabditis elegans* gene which is thought to be important in organism development.

This gene is expressed primarily in human synovial sarcoma tissue, bone marrow, and to a lesser extent in human brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, of bone, specifically synovial sarcoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the bone, connective tissues and possibly immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., synovial tissue, bone marrow, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another

tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

5 The tissue distribution and homology to *Caenorhabditis elegans* indicate that polynucleotides and polypeptides corresponding to this gene are useful as a diagnostic and/or therapeutic modality directed at the detection and/or treatment of connective tissue sarcomas or other related bone diseases.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 171

10 The translation product of this gene shares sequence homology with beta1-6GlcNAc transferase which is thought to be important in the transfer and metabolism of beta1-6, N-acetylglucosamine. This gene product has previously been shown to suppress melanoma lung metastasis in both syngeneic and nude mice, decreased invasiveness into the matrigel, and inhibition of cell attachment to collagen and laminin  
15 without affecting cell growth.

This gene is expressed primarily in human testes and prostate tissues, and to a lesser extent in kidney, medulla, and pancreas.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a  
20 biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer particularly melanoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at  
25 significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., testes and other reproductive tissue, prostate, kidney, pancreas, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard  
30 gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to beta1-6GlcNAc transferase indicate that the protein product of this clone is useful for the development of diagnostic and/or therapeutic modalities directed at the detection and/or treatment of cancer, the metastasis  
35 of malignant tissue or cells. Defects in this potentially secreted enzyme may play a role in metastasis.



**FEATURES OF PROTEIN ENCODED BY GENE NO: 172**

This gene is expressed primarily in fetal spleen and liver.

Therefore, polynucleotides and polypeptides of the invention are useful as  
5 reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions: immune disorders,  
Wilm's tumor disease, hepatic disorders, and hematopoietic disorders. Similarly,  
polypeptides and antibodies directed to these polypeptides are useful in providing  
immunological probes for differential identification of the tissue(s) or cell type(s). For a  
10 number of disorders of the above tissues or cells, particularly of the hematopoiesis and  
immune systems, expression of this gene at significantly higher or lower levels may be  
routinely detected in certain tissues and cell types (e.g., spleen and liver, and cancerous  
and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or  
spinal fluid) or another tissue or cell sample taken from an individual having such a  
15 disorder, relative to the standard gene expression level, i.e., the expression level in  
healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides  
corresponding to this gene are useful for the treatment and identification of fetal defects  
along with correcting diseases that affect hematopoiesis and the immune system.

20

**FEATURES OF PROTEIN ENCODED BY GENE NO: 173**

The translation product of this gene shares sequence homology with ret II  
oncogene which is thought to be important in Hirschsprung disease and many types of  
cancers.

25 This gene is expressed in multiple tissues including the lymphatic system, brain,  
and thyroid.

Therefore, polynucleotides and polypeptides of the invention are useful as  
reagents for identification of the tissue(s) or cell type(s) present in a biological sample  
and for diagnosis of diseases and conditions: Hirschsprung disease and multiple  
30 cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful  
in providing immunological probes for identification of the tissue(s) or cell type(s). For  
a number of disorders of the above tissues or cells, particularly of the immune and  
central nervous system, expression of this gene at significantly higher or lower levels  
may be routinely detected in certain tissues and cell types (e.g., lymphoid tissue,  
35 thyroid, and brain and other tissue of the nervous system, and cancerous and wounded  
tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or  
another tissue or cell sample taken from an individual having such a disorder, relative to

the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to ret II oncogene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of various cancers. It would also be useful for the diagnosis and treatment of Hirschsprung disease. Preferred polypeptides of the invention comprise the amino acid sequence: MEAQQVNEAESAREQLQXLHDQIAGQKASKQELETELERLKQEFHYIEEDLY RTKNTLQSRIKDRDEEIQKLRNQLTNKTLSSSSQSELENRLHQLTETLIQKQTMLESLSSTEKNSL VFQLERLEQQMNSASGSSSSNGSSINMSGIDNGEGTRLRNVPVLFNDTETNLAGMYGKVRKAAS  
 10 SIDQFSIRLGIFLRRYPIARVFIHYMALLHLWVMIVLLTYTPEM HHDQPYGK (SEQ ID NO: 642).

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 174

The translation product of this gene shares sequence homology with testis enhanced gene transcript which is thought to be important in regulation of human development.

This gene is expressed primarily in infant brain and to a lesser extent in a variety of other tissues and cell types, including the prostate, testes, monocytes, macrophages, dendritic cells, keratinocytes, and adipocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurological, developmental, immune and inflammation disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, prostate, testes and other reproductive tissue, blood cells, keratinocytes, and adipocytes, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to testis enhanced gene transcript indicate that the protein product of this clone is useful for diagnosis and treatment of disorders involving the developing brain and the immune system.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 175**

This gene is expressed primarily in prostate and to a lesser extent in various other tissues, including placenta.

5        Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancers, especially of the prostate. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for  
10 differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the prostate, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., prostate and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell  
15 sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for diagnosis and treatment of prostate disorders and cancer. It may also be useful for  
20 the diagnosis and treatment of endocrine disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 176**

The translation product of this gene shares sequence homology with  
25 *Sacchromyces cerevisiae* YNT20 gene which is thought to be important in mitochondrial function.

This gene is expressed at a particularly high level in muscle tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases related to such tissues and cell types  
30 including: muscle wasting diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neuromuscular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell  
35 types (e.g., muscle and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to the YNT20 gene indicate that this protein is useful for treatment and detection of neuromuscular diseases caused by loss of mitochondrial function. For example this gene or its protein product could be used in replacement therapy for such diseases.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 177**

This gene is expressed primarily in the brain and to a lesser extent in kidney, placenta, smooth muscle, heart and lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neuromuscular diseases, degenerative diseases of the central nervous system, and heart disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neuromuscular system, central nervous system, and heart, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, kidney, placenta, muscle, heart and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

This gene or its protein product could also be used for replacement therapy for the above mentioned diseases.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 178**

The translation product of this gene shares sequence homology with caldesmon which is thought to be important in the cellular response to changes in glucose levels.

This gene is expressed primarily in multiple tissues including brain and retina.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: central nervous system disorders and retinopathy. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for identification of the tissue(s) or cell

type(s). For a number of disorders of the above tissues or cells, particularly of the CNS disorders and retinopathy, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and retinal tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to caldesmon indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment of retinopathies.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 179**

The translation product of this gene shares sequence homology with mouse fibrosin protein which is thought to be important in regulation of fibrinogenesis in certain chronic inflammatory diseases.

This gene is expressed primarily in amniotic cells and breast tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of breast cancer and abnormal embryo development. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., amniotic cells, and mammary tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to fibrosin indicate that the protein product of this clone is useful for treatment of breast cancer. This gene or its protein product could be used in replacement therapy for breast cancer. In addition the protein product of this gene is useful in the treatment of chronic inflammatory diseases.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 180**

This gene is expressed several infant tissues including brain and liver and various adult tissues including brain, lung, liver, testes, and prostate.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, brain cancer, lung cancer, liver cancer and cancers of the reproductive system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, hepatic system, and reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, lung, liver, testes and other reproductive tissue, and prostate, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene product indicates that the protein product of this clone is involved in growth regulation and could be used as a growth factor or growth blocker in a variety of settings including treatment of cancers.

## **20 FEATURES OF PROTEIN ENCODED BY GENE NO: 181**

This gene is expressed primarily in activated monocytes and to a lesser extent in melanocytes and dendritic cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of immune system diseases and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, melanocytes, and dendritic cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone could be involved in growth regulation and could be used as a growth factor or growth blocker in a variety of settings including treatment of cancers.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 182**

This gene is expressed primarily in placenta and several tumors of various tissue origin and to a lesser extent in normal tissues including liver, lung, brain, and skin,

- 5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of cancers of all kinds. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders
- 10 of the above tissues or cells, particularly of the central nervous system, respiratory system and skin, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, lung, brain and other tissues of the nervous system, and skin, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or
- 15 cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- The high expression of this gene in multiple tumors indicates that the protein product of the clone may be involved in cell growth control and therefore would be
- 20 useful for treatment of certain cancers. Likewise molecules developed to block the activity of the protein product of this clone could be used to block its potential role in tumor growth promotion.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 183**

- 25 The translation product of this gene shares sequence homology with the mouse Ndr1 gene which is thought to be important in cancer progression.

This gene is expressed multiple cell types and tissues including brain, lung, kidney, bone marrow, liver, and spleen.

- Therefore, polynucleotides and polypeptides of the invention are useful as
- 30 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of all types of cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous, immune, and endocrine
- 35 systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, lung, kidney, bone marrow, liver and spleen, and cancerous and wounded

tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

5       The tissue distribution and homology to Ndr1 gene, which is thought to be involved in cancer progression, indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment of certain cancers. Likewise molecules developed to block the activity of the protein product of this clone could be used to block its potential role in tumor growth promotion.

10

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 184**

This gene is expressed primarily in early stage human brain and liver and to a lesser extent in several other fetal tissues.

15       Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: brain and liver cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the  
20       central nervous system and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, liver, and fetal tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder,  
25       relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The expression of this gene in embryonic tissues indicates that the protein could be involved in growth regulation and could be used as a growth factor or growth blocker in a variety of settings including treatment of cancers.

30

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 185**

This gene is expressed primarily in infant and embryonic brain.

35       Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of degenerative nervous system disorders and brain cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell



type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic tissue, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The expression of this gene in embryonic tissues indicates that the protein could be involved in growth regulation and could be used as a growth factor or growth blocker in a variety of settings including treatment of cancers.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 186**

This gene is expressed primarily in multiple tissues including placenta, fetal lung, fetal liver, and brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of all types of cancers including liver, brain and lung. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, pulmonary system, and hepatic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, lung, liver, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level; i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The expression of this gene in embryonic tissues indicates that the protein could be involved in growth regulation and could be used as a growth factor or growth blocker in a variety of settings including treatment of cancers.

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
1	HTTEZ21	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	11	582	1	582	177	177	313	1	18	19	22
1	HTTEZ21	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	197	1020	296	830	442	442	499	1	18	19	22
2	HBGBW52	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	12	465	1	465	81	81	314	1	30	31	128
2	HBGBW52	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	198	524	229	343		196	500	1	20	21	33
3	HCUFM41	97897 02/26/97 209043 05/15/97	ZAP Express	13	474	1	474	1	1	315	1	24	25	28
3	HCUFM41	97897 02/26/97 209043 05/15/97	ZAP Express	199	332	1	319	35	35	501	1	24	25	28

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
4	HCUFQ22	97897 02/26/97 209043 05/15/97	ZAP Express	14	314	1	298	122	122	316	1	34	35	64
5	HCUFV01	97897 02/26/97 209043 05/15/97	ZAP Express	15	613	1	613	30	30	317	1	18	19	21
6	HCUGA50	97897 02/26/97 209043 05/15/97	ZAP Express	16	356	1	356	239	239	318	1	22	23	39
7	HCUIM14	97897 02/26/97 209043 05/15/97	ZAP Express	17	414	185	414	278	278	319	1	26	27	33
8	HLD0U93	97897 02/26/97 209043 05/15/97	pCMVSPORT 3.0	18	469	1	469	77	77	320	1	44	45	88
9	HEIAX07	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	19	550	1	550	129	129	321	1	21	22	23

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	5' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
9	HEIAX07	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	200	376	9	376		1	502	1	8	9	15
10	HSAXR76	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	20	741	55	741	190	190	322	1			27
11	HNGJJ68	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	21	991	1	991	62	62	323	1	30	31	64
11	HNGJJ68	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	201	1192	253	1137		409	503	1			19
12	HCFW04	97897 02/26/97 209043 05/15/97	pSport1	22	653	1	653	64	64	324	1	30	31	196
12	HCFW04	97897 02/26/97 209043 05/15/97	pSport1	202	589	1	513	109	109	504	1			29

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
13	HLMV65	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	23	1486	596	1418	102	102	325	1	54	55	252
13	HLMV65	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	203	847	1	839	87	87	505	1	30	31	75
13	HLMV65	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	204	852	75	850		690	506	1			10
13	HTXEF04	209235 09/04/97	Uni-ZAP XR	205	1354	54	1354	100	100	507	1	33	34	207
14	HPMFD84	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	24	2323	1017	2059	1242	1242	326	1	21	22	68
14	HPMFD84	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	206	1378	113	1226	303	303	508	1	25	26	36
15	HE6DB26	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	25	683	1	683	304	304	327	1	30	31	84

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
15	HE6DB26	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	207	1166	281	884	567	567	509	1	18	19	19
16	HHFFL33	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	26	2036	14	1959	214	214	328	1	20	21	36
17	HODBD33	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	27	717	1	717	70	70	329	1	30	31	63
17	HODBD33	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	208	697	2	697	33	33	510	1	31	32	32
18	HMDAE90	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	28	495	1	495	39	39	330	1	24	25	35
19	HOUAW01	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	29	556	1	556	116	116	331	1	19	20	23

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
20	HBJAE44	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	30	434	1	434	78	78	332	1	35	36	40
21	HCFME41	97897 02/26/97 209043 05/15/97	pSport1	31	715	1	715	87	87	333	1	30	31	111
21	HCFME41	97897 02/26/97 209043 05/15/97	pSport1	209	932	274	932	387	387	511	1	27	28	28
22	HOGCO71	97897 02/26/97 209043 05/15/97	pCMVSPORT 2.0	32	486	1	486	137	137	334	1	21	22	106
23	HOSEX08	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	33	725	1	725	436	436	335	1	30	31	50
23	HOSEX08	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	210	661	1	647	81	81	512	1	25	26	26

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
24	HSKNJ72	97897 02/26/97 209043 05/15/97	pBluescript	34	437	1	437	85	85	336	1	30	31	48
25	HEBEB69	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	35	943	1	943	196	196	337	1	30	31	41
25	HEBEB69	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	211	592	1	534	72	72	513	1	24	25	33
26	HE6EH18	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	36	604	1	604	375	375	338	1	20	21	76
26	HE6EH18	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	212	938	1	509		17	514	1	30	31	47
27	HSAUZ47	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	37	349	1	349		201	339	1	20	21	31



Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
28	HSSDM73	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	38	672	1	672	22	22	340	1	38	39	42
29	HBMVK68	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	39	1908	135	1908	309	309	341	1	20	21	26
30	HMKDC66	97898 02/26/97 209044 05/15/97	pSport1	40	458	93	458	147	147	342	1	24	25	26
31	HMKCU94	97898 02/26/97 209044 05/15/97	pSport1	41	1153	500	1153	427	427	343	1	30	31	157
31	HMKCU94	97898 02/26/97 209044 05/15/97	pSport1	213	1079	502	896	739	739	515	1	23	24	43
32	HRDEW41	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	42	1983	1092	1983	27	27	344	1	11	12	520

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
32	HRDEW41	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	214	3791	2757	3357		2030	516	1			3
33	HTOJN06	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	43	1406	1	695		19	345	1	19	20	39
34	HBGDA21	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	44	1391	851	1153	74	74	346	1	30	31	234
34	HBGDA21	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	215	1334	822	1036		638	517	1	18	19	174
35	HFGAK75	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	45	1569	768	1569	14	14	347	1	19	20	169
35	HFGAK75	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	216	1511	770	1404	844	844	518	1	32	33	43

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
36	HHPBD40	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	46	1924	1	1681	62	62	348	1	19	20	43
37	HOVCL83	97898 02/26/97 209044 05/15/97	pSport1	47	475	252	396	141	141	349	1	37	38	78
38	HBCAY62	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	48	346	1	346	61	61	350	1	19	20	24
39	HBICM48	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	49	1366	882	1300	177	177	351	1	30	31	274
39	HBICM48	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	217	642	192	581		448	519	1			13
40	HLTCL35	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	50	1405	110	1404	61	61	352	1	30	31	47

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
40	HLTCL35	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	218	1241	1	1241	172	172	520	1	21	22	30
41	HLHCK50	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	51	504	207	485	222	222	353	1			3
42	HRSAN45	97899 02/26/97 209045 05/15/97	ZAP Express	52	777	1	214	113	113	354	1	24	25	52
43	HSNBB14	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	53	602	1	419	41	41	355	1	59	60	132
43	HSNBB14	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	219	1080	186	686	399	399	521	1	26	27	47
44	HMABL38	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	54	1749	222	1749	166	166	356	1	30	31	204

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
44	HMABL38	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	220	1258	149	1190	254	254	522	1	18	19	26
45	HSKDK47	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	55	1896	596	1614	650	650	357	1	33	34	47
46	HOSFH03	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	56	1753	555	1753	414	414	358	1	18	19	73
46	HOSFH03	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	221	1693	554	1693		526	523	1	25	26	58
47	HOGAV75	97899 02/26/97 209045 05/15/97	pCMVSPORT 2.0	57	1220	690	1024	128	128	359	1	30	31	102
47	HOGAV75	97899 02/26/97 209045 05/15/97	pCMVSPORT 2.0	222	1196	712	1163		1097	524	1			19

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
48	HFCA174	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	58	1049	362	1049	335	335	360	1	33	34	48
49	HAGBI17	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	59	1776	854	1737	189	189	361	1	30	31	179
49	HAGBI17	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	223	1791	979	1791	1164	1164	525	1	18	19	40
50	HLFBC91	97899 02/26/97 209045 05/15/97	pBluescript SK-	60	443	1	443	164	164	362	1	21	22	25
51	HPRCA31	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	61	2888	1909	2888	90	90	363	1	30	31	224
51	HPRCA31	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	224	2517	1597	2517	1953	1953	526	1	18	19	57

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
52	HPRCE95	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	62	1851	1568	1736	139	139	364	1	30	31	349
52	HPRCE95	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	225	2424	299	2309	530	530	527	1	17	18	21
53	HHTLC66	97899 02/26/97 209045 05/15/97	ZAP Express	63	3542	883	3492	964	964	365	1	25	26	467
54	HMADI02	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	64	883	237	883	229	229	366	1	30	31	152
54	HMADI02	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	226	1080	242	1033	436	436	528	1	24	25	39
55	HPRCU93	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	65	1541	1	1541	236	236	367	1	30	31	373

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
55	HPRCU93	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	227	1336	4	1336	946	946	529	1	25	26	128
56	HSAXS65	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	66	732	41	698	163	163	368	1	18	19	83
56	HSAXS65	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	228	2043	1133	1756	1262	1262	530	1	20	21	82
57	HKTAG35	209011 04/28/97	Uni-ZAP XR	67	629	1	629	264	264	369	1			21
57	HMEFX42	97899 02/26/97 209045 05/15/97	Lambda ZAP II	229	540	25	536	227	227	531	1			20
58	HHFHN61	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	68	1751	375	1751	95	95	370	1	19	20	227
59	HCWEF90	97899 02/26/97 209045 05/15/97	ZAP Express	69	508	1	508	22	22	371	1	30	31	79



Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of AA of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
59	HCWEF90	97899 02/26/97 209045 05/15/97	ZAP Express	230	448	9	448		1	532	1	22	23	75
60	HHGCM20	97899 02/26/97 209045 05/15/97	Lambda ZAP II	70	245	1	245	93	93	372	1	1	2	51
61	HFRAU10	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	71	361	1	361	1	1	373	1	30	31	61
61	HFRAU10	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	231	407	1	407	210	210	533	1	17	18	60
62	HATDT67	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	72	713	8	713	169	169	374	1	30	31	40
62	HATDT67	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	232	830	190	580	329	329	534	1	28	29	39

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
63	HOUBG93	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	73	862	1	862	67	67	375	1	30	31	44
63	HOUBG93	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	233	932	138	905	287	287	535	1			2
64	HMWEX24	97900 02/26/97 209046 05/15/97	Uni-Zap XR	74	4602	4162	4525	730	730	376	1	30	31	203
64	HMWEX24	97900 02/26/97 209046 05/15/97	Uni-Zap XR	234	2786	2406	2739	2577	2577	536	1	22	23	36
65	HSGBA84	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	75	1255	1	1195	112	112	377	1	28	29	29
66	HTOCD52	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	76	475	1	475	13	13	378	1	30	31	136

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
66	HTOCD52	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	235	458	1	458	26	537	1			14
67	HTGCP16	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	77	465	25	299	74	379	1	33	34	41
68	HKIXR69	97900 02/26/97 209046 05/15/97	pBluescript	78	1907	1627	1730	26	380	1	30	31	468
68	HKIXR69	97900 02/26/97 209046 05/15/97	pBluescript	236	591	1	444	251	538	1			18
69	HETGJ09	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	79	1168	136	1168	267	381	1	20	21	29
70	HOBNC61	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	80	1285	132	1285	292	382	1	27	28	29

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT3' NT of Clone Seq.	5' NT of Clone Seq.	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
71	HFFAH94	97900 02/26/97 209046 05/15/97	Lambda ZAP II	81	1290	768	1054	701	383	1	21	22	138
72	HBIAB95	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	82	684	1	684	119	384	1	30	31	74
73	HSQEL25	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	83	2024	1609	1953	200	385	1	30	31	521
73	HSQEL25	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	237	1286	391	959	1204	539	1	9	10	11
74	HEBEG68	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	84	931	14	537	85	386	1	25	26	137
75	HBIAB39	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	85	825	59	802	66	387	1	30	31	186

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
75	HBIAB39	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	238	734	1	734	1	540	1	37	38	108
75	HBIAB39	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	239	809	80	794	294	541	1	15	16	106
76	HTXDU73	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	86	1238	36	918	17	388	1			1
77	HOEAS24	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	87	1460	9	1458	166	389	1	53	54	299
77	HOEAS24	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	240	2201	841	2080	507	542	1	43	44	136
77	HOEAS24	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	241	1661	311	1520	390	543	1	35	36	424

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
78	HTEIY30	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	88	1395	567	1395	639	639	390	1	36	37	49
79	HSKNE46	97900 02/26/97 209046 05/15/97	pBluescript	89	1186	352	1186	540	540	391	1	49	50	61
79	HSKNE46	97900 02/26/97 209046 05/15/97	pBluescript	242	1146	329	1146	564	564	544	1	21	22	39
80	HPMFL27	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	90	1821	1203	1614	1503	1503	392	1	30	31	79
81	HMWDN32	97900 02/26/97 209046 05/15/97	Uni-Zap XR	91	862	253	862	359	359	393	1	32	33	36
82	HPRAX55	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	92	696	349	696	98	98	394	1	30	31	180

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	5' NT of Clone Seq.	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
82	HPRAX55	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	243	1350	265	1230	348	545	1	32	33	58
83	HHFFW36	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	93	1886	1	1759	197	395	1			21
84	HE2PL77	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	94	1774	742	1772	785	396	1	21	22	60
85	HSDFV29	209076 05/22/97	Uni-ZAP XR	95	2503	1	1648	206	397	1	32	33	152
85	HCQAV53	97901 02/26/97 209047 05/15/97	Lambda ZAP II	244	1529	72	911	191	546	1	20	21	33
86	HTPEG42	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	96	2801	418	2801	234	398	1	30	31	480
86	HTPEG42	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	245	1537	1	1537	125	547	1	21	22	367

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
87	HLHDR57	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	97	1631	916	1631	1	399	1	1	2	423
88	HAUAV32	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	98	504	26	504	197	400	1	23	24	78
88	HAUAV32	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	246	506	1	499	183	548	1	32	33	77
89	HNEBI60	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	99	1416	145	1416	456	401	1	18	19	74
89	HNEBI60	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	247	1348	84	1348	363	549	1	21	22	47
90	HSHCI16	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	100	2847	1	2847		402	1			20



Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	5' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
91	HTSEL31	97901 02/26/97 209047 05/15/97	pBluescript	101	1394	608	1346	602	602	403	1	23	24	87
92	HAUBL57	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	102	794	1	794	518	518	404	1	30	31	92
92	HAUBL57	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	248	1766	42	1766	356	356	550	1	30	31	168
92	HAUBL57	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	249	2664	47	1708		147	551	1	18	19	124
93	HODASS9	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	103	1544	898	1531	975	975	405	1			21
94	HE6CT48	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	104	871	106	871	248	248	406	1	34	35	174

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
94	HE6CT48	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	250	865	97	865	258	258	552	1	19	20	177
95	HMDAA61	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	105	404	1	404	16	16	407	1	21	22	64
95	HMDAA61	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	251	2082	852	2074	829	829	553	1	22	23	72
96	HAQBK61	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	106	1542	506	1542	122	122	408	1	51	52	280
96	HAQBK61	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	252	1482	508	1482	633	633	554	1	15	16	45
96	HCUHB01	209215 08/21/97	ZAP Express	253	834	1	834	82	82	555	1	40	41	251
97	HAQBF73	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	107	2327	1528	2327	465	465	409	1	30	31	284

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
97	HAQBF73	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	254	1508	885	1508		988	556	1			19
98	HAQBT94	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	108	1062	157	1062	172	172	410	1	28	29	187
99	HETHE07	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	109	2539	275	2501	903	903	411	1	30	31	237
99	HETHE07	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	255	2514	592	2431	176	176	557	1	30	31	217
99	HETHE07	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	256	2357	465	2288		1151	558	1	12	13	82
100	HLQAB52	97901 02/26/97 209047 05/15/97	Lambda ZAP II	110	1751	969	1751	4	4	412	1	46	47	192

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
100	HLQAB52	97901 02/26/97 209047 05/15/97	Lambda ZAP II	257	689	218	655	314	314	559	1	18	19	95
100	HEONN58	209119 06/12/97	pSport1	258	2377	5	2377	25	25	560	1	28	29	54
101	HCRAM28	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	111	1117	1	1117		1	413	1	19	20	21
101	HIBEK16	209627 02/12/98	Other	259	1193	69	1135	242	242	561	1	24	25	108
102	HE2BG03	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	112	1313	128	1313	271	271	414	1	30	31	51
102	HE2BG03	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	260	1262	26	1262	35	35	562	1	35	36	50
103	HEBDJ82	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	113	1654	553	1654	709	709	415	1			32

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
104	HCUBC79	97901 02/26/97 209047 05/15/97	ZAP Express	114	1171	540	1171	337	337	416	1	30	31	163
104	HCUBC79	97901 02/26/97 209047 05/15/97	ZAP Express	261	1179	626	1161	335	335	563	1	30	31	253
104	HCUBC79	97901 02/26/97 209047 05/15/97	ZAP Express	262	1162	629	1131	942	942	564	1			18
105	HSVAF07	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	115	842	373	800	100	100	417	1	65	66	174
105	HSVAF07	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	263	735	290	735			565	1			
105	HSVAF07	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	264	783	416	783		413	566	1	33	34	73

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	5' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
106	HT3AM65	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	116	1640	187	1470	581	581	418	1	30	31	50
106	HT3AM65	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	265	1638	301	1405	119	119	567	1	30	31	263
106	HT3AM65	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	266	1455	148	1188	438	438	568	1	24	25	70
107	HE6DK18	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	117	952	418	906	499	499	419	1	28	29	120
108	HEBEK93	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	118	1256	21	1079	301	301	420	1	30	31	159
108	HEBEK93	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	267	1086	25	1050	227	227	569	1	23	24	34

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
109	HJPCM10	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	119	1143	171	1051	175	421	1	50	51	154
109	HJPCM10	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	268	1003	21	1003	115	570	1	34	35	104
109	HJPCM10	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	269	1234	174	1015	232	571	1	27	28	132
110	HSXBL78	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	120	1782	1	1720	138	422	1	32	33	204
111	HOEAW81	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	121	610	18	609	50	423	1	30	31	67
111	HOEAW81	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	270	574	1	566	337	572	1	27	28	32

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
112	HOEAP41	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	122	526	185	375	143	424	1	21	22	25
113	HEAAR60	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	123	2081	1179	1976	48	425	1	30	31	299
113	HEAAR60	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	271	1731	889	1626	886	573	1	18	19	28
114	HTXGS75	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	124	1717	764	1640	76	426	1			13
115	HOVBA03	97902 02/26/97 209048 05/15/97	pSport1	125	804	1	804	145	427	1	15	16	198
115	HOVBA03	97902 02/26/97 209048 05/15/97	pSport1	272	1320	77	637	280	574	1	22	23	40



Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
116	HGBGK76	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	126	431	1	431	73	73	428	1	38	39	47
116	HGBGK76	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	273	515	1	515	43	43	575	1	20	21	30
117	HBMUW78	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	127	3752	3465	3752	748	748	429	1	30	31	370
117	HBMUW78	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	274	2995	2738	2995	2777	2777	576	1	18	19	29
118	HASAS24	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	128	1144	669	1144	896	896	430	1			30
119	HSIDN55	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	129	1830	1234	1830	1265	1265	431	1			24

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
120	HGBGZ64	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	130	1864	1505	1741	1578	1578	432	1	37	38	53
121	H6EBJ64	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	131	2041	1	1214	46	46	433	1	35	36	176
121	H6EBJ64	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	275	1990	8	1128	71	71	577	1	16	17	92
122	HOECP43	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	132	2012	853	1986	1127	1127	434	1	22	23	77
123	H2CBV31	97902 02/26/97 209048 05/15/97	pBluescript SK-	133	1669	670	1632	962	962	435	1	25	26	32
124	HPCAD23	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	134	1565	281	1565	274	274	436	1	25	26	30

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
125	HSPAG15	97902 02/26/97 209048 05/15/97	pSport1	135	2007	1101	2007	1124	1124	437	1	39	40	69
126	HELGH31	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	136	1291	1	1180	107	107	438	1			19
127	HUSHH48	97902 02/26/97 209048 05/15/97	Lambda ZAP II	137	1906	1	1906	184	184	439	1	30	31	43
127	HUSHH48	97902 02/26/97 209048 05/15/97	Lambda ZAP II	276	2436	572	2436	726	726	578	1	30	31	42
128	HLYAU95	97902 02/26/97 209048 05/15/97	pSport1	138	1935	1044	1794	1183	1183	440	1	18	19	33
129	HHSCV65	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	139	1446	572	1347	585	585	441	1	25	26	53

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	5' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
130	HTTAD57	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	140	1109	639	1109	676	676	442	1	24	25	64
131	HEBGA37	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	141	497	9	497	95	95	443	1			34
132	HEBFU93	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	142	269	1	269	1	1	444	1	30	31	89
132	HEBFU93	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	277	782	408	781		571	579	1	31	32	70
133	HSGSC60	97902 02/26/97 209048 05/15/97	Lambda ZAP II	143	1269	55	1262	55	55	445	1	25	26	350
134	HPMGD24	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	144	1944	97	1871	306	306	446	1	16	17	49

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
135	HPTVC60	97902 02/26/97 209048 05/15/97	pBluescript	145	1021	526	1021	74	447	1	30	31	278
135	HPTVC60	97902 02/26/97 209048 05/15/97	pBluescript	278	961	524	961	545	580	1	23	24	110
136	HSKNE18	97902 02/26/97 209048 05/15/97	pBluescript	146	1285	5	1285	116	448	1	30	31	199
136	HSKNE18	97902 02/26/97 209048 05/15/97	pBluescript	279	1228	9	1228	324	581	1	26	27	30
137	HMWIF35	97902 02/26/97 209048 05/15/97	Uni-Zap XR	147	1386	169	1272	165	449	1	30	31	258
137	HMWIF35	97902 02/26/97 209048 05/15/97	Uni-Zap XR	280	1327	169	1208	160	582	1	23	24	71

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
138	HMWGI25	97902 02/26/97 209048 05/15/97	Uni-Zap XR	148	2098	721	2044	784	784	450	1	18	19	87
139	HSKGF03	97902 02/26/97 209048 05/15/97	pBluescript	149	1847	1689	1847	241	241	451	1	33	34	315
139	HSKGF03	97902 02/26/97 209048 05/15/97	pBluescript	281	799	1	799		243	583	1	12	13	47
140	HMSKE75	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	150	1569	113	1517	417	417	452	1	21	22	52
141	HCMSH30	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	151	1540	538	1540	48	48	453	1	30	31	383
141	HCMSH30	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	282	2196	270	2196	294	294	584	1	32	33	39

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
142	HTWCB92	97902 02/26/97 209048 05/15/97	pSport1	152	1719	690	1575	6	6	454	1	52	53	186
143	HBMDM46	97902 02/26/97 209048 05/15/97	pBluescript	153	863	1	863	195	195	455	1	26	27	163
143	HBMDM46	97902 02/26/97 209048 05/15/97	pBluescript	283	1185	277	1166	621	621	585	1			19
144	HFAMG13	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	154	1101	1	512	40	40	456	1	21	22	46
145	HFXHL79	97903 02/26/97 209049 05/15/97	Lambda ZAP II	155	2031	669	2031	411	411	457	1	23	24	105
145	HFXHL79	97903 02/26/97 209049 05/15/97	Lambda ZAP II	284	1634	615	1485	878	878	586	1	20	21	23

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
146	HSNAK17	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	156	1981	1458	1809	1592	458	1	23	24	70
146	HSNAK17	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	285	1795	1458	1749	1562	587	1	33	34	69
147	HCFBC03	97903 02/26/97 209049 05/15/97	pSport1	157	915	45	912	22	459	1	22	23	155
147	HCFBC03	97903 02/26/97 209049 05/15/97	pSport1	286	858	46	858	224	588	1	30	31	77
147	HSJAP03	209139 07/03/97	Uni-ZAP XR	287	915	1	915	22	589	1	22	23	155
148	HSKGO26	97903 02/26/97 209049 05/15/97	pBluescript	158	2117	51	1422	32	460	1	23	24	332
149	HCQAV96	97903 02/26/97 209049 05/15/97	Lambda ZAP II	159	2395	1509	2382	1440	461	1			5



Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
150	HSHCC16	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	160	2120	1223	2108	1416	462	1			14
151	HTLEF62	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	161	900	482	900	46	463	1	30	31	285
151	HTLEF62	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	288	1517	783	1517	1062	590	1			24
152	HTLAD94	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	162	1003	1	1003	288	464	1	30	31	80
152	HTLAD94	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	289	3865	217	1195	281	591	1	16	17	38
153	HTSFQ12	97903 02/26/97 209049 05/15/97	pBluescript	163	2196	1607	2180	1611	465	1	30	31	47

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
154	HE6FL83	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	164	1945	271	1840	299	299	466	1	63	64	96
154	HE6FL83	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	290	1910	279	1818	355	355	592	1	39	40	69
155	HTXFJ55	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	165	2933	489	2871	258	258	467	1	30	31	399
155	HTXFJ55	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	291	3276	486	2838		525	593	1	45	46	308
156	HJPCJ76	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	166	2243	343	2221		341	468	1			1
157	HLTED27	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	167	1816	1130	1816	284	284	469	1	31	32	273

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
157	HLTED27	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	292	1695	1098	1548	1306	1306	594	1			22
158	HMKBA64	97903 02/26/97 209049 05/15/97	pSport1	168	945	1	787	208	208	470	1	18	19	192
159	HNFI24	97903 02/26/97 209049 05/15/97	pBluescript	169	902	46	816	19	19	471	1	26	27	234
160	HCELB21	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	170	1883	798	1869	1001	1001	472	1	45	46	105
160	HCELB21	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	293	1501	438	1501	510	510	595	1			24
161	HAWBA28	97903 02/26/97 209049 05/15/97	pBluescript SK-	171	2100	1642	2100	1722	1722	473	1	23	24	32

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
162	HSAAS44	97903 02/26/97 209049 05/15/97	pBluescript SK-	172	1930	187	1930	65	474	1	30	31	571
162	HSAAS44	97903 02/26/97 209049 05/15/97	pBluescript SK-	294	2683	183	2683	431	596	1			24
163	HAFAL73	97903 02/26/97 209049 05/15/97	pBluescript SK-	173	1509	962	1451	122	475	1	30	31	312
163	HAFAL73	97903 02/26/97 209049 05/15/97	pBluescript SK-	295	1454	961	1420	976	597	1			1
164	HSAWF26	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	174	3173	2197	2972	51	476	1	21	22	329
164	HSAWF26	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	296	828	52	828	305	598	1			8

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
165	HEAAL31	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	175	991	374	970	60	60	477	1	24	25	178
165	HEAAL31	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	297	2416	1387	2413	1473	1473	599	1	18	19	25
166	HFKFX55	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	176	1290	499	1290		688	478	1	25	26	52
167	H2LAO11	97903 02/26/97 209049 05/15/97	pBluescript SK-	177	2290	1	2290	173	173	479	1	22	23	62
168	HPFDZ95	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	178	549	1	549	11	11	480	1	21	22	27
168	HPFDZ95	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	298	545	1	545	17	17	600	1	21	22	27

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
169	HPTTU11	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	179	1509	294	1352	92	92	481	1	30	31	339
169	HPTTU11	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	299	1530	385	1530	562	562	601	1	23	24	61
170	HCFAE79	97904 02/26/97 209050 05/15/97	pSport1	180	1316	985	1250	995	995	482	1	26	27	32
171	HTEDJ34	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	181	777	1	777	51	51	483	1	30	31	48
171	HTEDJ34	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	300	997	244	997	300	300	602	1	23	24	29
172	HODCW06	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	182	791	1	791	14	14	484	1	29	30	38

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
173	HFTAR26	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	183	1405	346	1405	575	575	485	1	20	21	61
174	H2MBF44	97904 02/26/97 209050 05/15/97	pBluescript SK-	184	1596	75	1596	131	131	486	1	24	25	346
174	H2MBF44	97904 02/26/97 209050 05/15/97	pBluescript SK-	301	2345	75	2345	233	233	603	1	56	57	69
175	HE8BI92	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	185	2293	355	2288	67	67	487	1	30	31	237
175	HE8BI92	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	302	2369	2	1946		60	604	1	9	10	24
176	HFTBR48	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	186	1212	462	1180	257	257	488	1	30	31	200

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
176	HFTBR48	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	303	1181	424	1149	663	663	605	1	23	24	35
177	HE9CM64	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	187	1605	770	1554	166	166	489	1	30	31	351
177	HE9CM64	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	304	1537	719	1515		787	606	1	43	44	130
178	HATAV51	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	188	1516	960	1516	8	8	490	1	30	31	265
178	HATAV51	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	305	1493	1	1261	54	54	607	1	18	19	23
179	HAQAF27	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	189	681	287	681		401	491	1			25



Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
180	HCEEK08	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	190	1014	703	1014	360	360	492	1	30	31	159
180	HCEEK08	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	306	577	1	577		175	608	1			6
181	HAFUI8	97904 02/26/97 209050 05/15/97	pBluescript SK-	191	2779	2207	2630	1153	1153	493	1	30	31	279
181	HAFUI8	97904 02/26/97 209050 05/15/97	pBluescript SK-	307	2860	163	2860	21	21	609	1	30	31	232
181	HAFUI8	97904 02/26/97 209050 05/15/97	pBluescript SK-	308	876	275	876	302	302	610	1	32	33	34
182	HETBY74	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	192	1923	30	1923	45	45	494	1	33	34	193

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
183	HTOAF35	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	193	2346	1160	2286	178	178	495	1	30	31	205
183	HTOAF35	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	309	2025	840	2025	971	971	611	1	18	19	21
184	HCRBX32	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	194	3054	2004	3054	434	434	496	1	11	12	147
184	HCRBX32	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	310	3026	1966	3026	2131	2131	612	1			9
185	HEGBB80	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	195	907	152	907	297	297	497	1	30	31	64
185	HEGBB80	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	311	712	67	712	107	107	613	1	18	19	29

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
186	HFAMH74	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	196	1290	84	809	225	225	498	1	30	31	94
186	HFAMH74	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	312	1289	785	1289	927	927	614	1	28	29	30

Table 1 summarizes the information corresponding to each "Gene No." described above. The nucleotide sequence identified as "NT SEQ ID NO:X" was assembled from partially homologous ("overlapping") sequences obtained from the "cDNA clone ID" identified in Table 1 and, in some cases, from additional related DNA clones. The overlapping sequences were assembled into a single contiguous sequence of high redundancy (usually three to five overlapping sequences at each nucleotide position), resulting in a final sequence identified as SEQ ID NO:X.

The cDNA Clone ID was deposited on the date and given the corresponding deposit number listed in "ATCC Deposit No:Z and Date." Some of the deposits contain multiple different clones corresponding to the same gene. "Vector" refers to the type of vector contained in the cDNA Clone ID.

"Total NT Seq." refers to the total number of nucleotides in the contig identified by "Gene No." The deposited clone may contain all or most of these sequences, reflected by the nucleotide position indicated as "5' NT of Clone Seq." and the "3' NT of Clone Seq." of SEQ ID NO:X. The nucleotide position of SEQ ID NO:X of the putative start codon (methionine) is identified as "5' NT of Start Codon." Similarly, the nucleotide position of SEQ ID NO:X of the predicted signal sequence is identified as "5' NT of First AA of Signal Pep."

The translated amino acid sequence, beginning with the methionine, is identified as "AA SEQ ID NO:Y," although other reading frames can also be easily translated using known molecular biology techniques. The polypeptides produced by these alternative open reading frames are specifically contemplated by the present invention.

The first and last amino acid position of SEQ ID NO:Y of the predicted signal peptide is identified as "First AA of Sig Pep" and "Last AA of Sig Pep." The predicted first amino acid position of SEQ ID NO:Y of the secreted portion is identified as "Predicted First AA of Secreted Portion." Finally, the amino acid position of SEQ ID NO:Y of the last amino acid in the open reading frame is identified as "Last AA of ORF."

SEQ ID NO:X and the translated SEQ ID NO:Y are sufficiently accurate and otherwise suitable for a variety of uses well known in the art and described further below. For instance, SEQ ID NO:X is useful for designing nucleic acid hybridization probes that will detect nucleic acid sequences contained in SEQ ID NO:X or the cDNA contained in the deposited clone. These probes will also hybridize to nucleic acid molecules in biological samples, thereby enabling a variety of forensic and diagnostic methods of the invention. Similarly, polypeptides identified from SEQ ID NO:Y may be used to generate antibodies which bind specifically to the secreted proteins encoded by the cDNA clones identified in Table 1.

Nevertheless, DNA sequences generated by sequencing reactions can contain sequencing errors. The errors exist as misidentified nucleotides, or as insertions or deletions of nucleotides in the generated DNA sequence. The erroneously inserted or deleted nucleotides cause frame shifts in the reading frames of the predicted amino acid sequence. In these cases, the predicted amino acid sequence diverges from the actual amino acid sequence, even though the generated DNA sequence may be greater than 99.9% identical to the actual DNA sequence (for example, one base insertion or deletion in an open reading frame of over 1000 bases).

Accordingly, for those applications requiring precision in the nucleotide sequence or the amino acid sequence, the present invention provides not only the generated nucleotide sequence identified as SEQ ID NO:X and the predicted translated amino acid sequence identified as SEQ ID NO:Y, but also a sample of plasmid DNA containing a human cDNA of the invention deposited with the ATCC, as set forth in Table 1. The nucleotide sequence of each deposited clone can readily be determined by sequencing the deposited clone in accordance with known methods. The predicted amino acid sequence can then be verified from such deposits. Moreover, the amino acid sequence of the protein encoded by a particular clone can also be directly determined by peptide sequencing or by expressing the protein in a suitable host cell containing the deposited human cDNA, collecting the protein, and determining its sequence.

The present invention also relates to the genes corresponding to SEQ ID NO:X, SEQ ID NO:Y, or the deposited clone. The corresponding gene can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include preparing probes or primers from the disclosed sequence and identifying or amplifying the corresponding gene from appropriate sources of genomic material.

Also provided in the present invention are species homologs. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source for the desired homologue.

The polypeptides of the invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

The polypeptides may be in the form of the secreted protein, including the mature form, or may be a part of a larger protein, such as a fusion protein (see below).

It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification, such as multiple histidine residues, or an additional sequence for stability during recombinant production.

5       The polypeptides of the present invention are preferably provided in an isolated form, and preferably are substantially purified. A recombinantly produced version of a polypeptide, including the secreted polypeptide, can be substantially purified by the one-step method described in Smith and Johnson, *Gene* 67:31-40 (1988). Polypeptides of the invention also can be purified from natural or recombinant sources  
10       using antibodies of the invention raised against the secreted protein in methods which are well known in the art.

### Signal Sequences

15       Methods for predicting whether a protein has a signal sequence, as well as the cleavage point for that sequence, are available. For instance, the method of McGeoch, *Virus Res.* 3:271-286 (1985), uses the information from a short N-terminal charged region and a subsequent uncharged region of the complete (uncleaved) protein. The method of von Heinje, *Nucleic Acids Res.* 14:4683-4690 (1986) uses the information from the residues surrounding the cleavage site, typically residues -13 to +2; where +1  
20       indicates the amino terminus of the secreted protein. The accuracy of predicting the cleavage points of known mammalian secretory proteins for each of these methods is in the range of 75-80%. (von Heinje, *supra*.) However, the two methods do not always produce the same predicted cleavage point(s) for a given protein.

25       In the present case, the deduced amino acid sequence of the secreted polypeptide was analyzed by a computer program called SignalP (Henrik Nielsen et al., *Protein Engineering* 10:1-6 (1997)), which predicts the cellular location of a protein based on the amino acid sequence. As part of this computational prediction of localization, the methods of McGeoch and von Heinje are incorporated. The analysis of the amino acid sequences of the secreted proteins described herein by this program provided the results  
30       shown in Table 1.

As one of ordinary skill would appreciate, however, cleavage sites sometimes vary from organism to organism and cannot be predicted with absolute certainty. Accordingly, the present invention provides secreted polypeptides having a sequence shown in SEQ ID NO:Y which have an N-terminus beginning within 5 residues (i.e., +  
35       or - 5 residues) of the predicted cleavage point. Similarly, it is also recognized that in some cases, cleavage of the signal sequence from a secreted protein is not entirely

uniform, resulting in more than one secreted species. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

Moreover, the signal sequence identified by the above analysis may not necessarily predict the naturally occurring signal sequence. For example, the naturally occurring signal sequence may be further upstream from the predicted signal sequence. However, it is likely that the predicted signal sequence will be capable of directing the secreted protein to the ER. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

#### 10 **Polynucleotide and Polypeptide Variants**

"Variant" refers to a polynucleotide or polypeptide differing from the polynucleotide or polypeptide of the present invention, but retaining essential properties thereof. Generally, variants are overall closely similar, and, in many regions, identical to the polynucleotide or polypeptide of the present invention.

15 "Identity" per se has an art-recognized meaning and can be calculated using published techniques. (See, e.g.: (COMPUTATIONAL MOLECULAR BIOLOGY, Lesk, A.M., ed., Oxford University Press, New York, (1988); BIOCOMPUTING: INFORMATICS AND GENOME PROJECTS, Smith, D.W., ed., Academic Press, New York, (1993); COMPUTER ANALYSIS OF SEQUENCE DATA, PART I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, (1994);  
20 SEQUENCE ANALYSIS IN MOLECULAR BIOLOGY, von Heinje, G., Academic Press, (1987); and SEQUENCE ANALYSIS PRIMER, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, (1991).) While there exists a number of methods to measure identity between two polynucleotide or polypeptide sequences, the term "identity" is well known to skilled artisans. (Carillo, H., and Lipton, D., SIAM J Applied Math 48:1073 (1988).) Methods commonly employed to determine identity or similarity between two sequences include, but are not limited to, those disclosed in "Guide to Huge Computers," Martin J. Bishop, ed., Academic Press, San Diego, (1994), and Carillo, H., and Lipton, D., SIAM J Applied Math 48:1073 (1988).

30 Methods for aligning polynucleotides or polypeptides are codified in computer programs, including the GCG program package (Devereux, J., et al., Nucleic Acids Research (1984) 12(1):387 (1984)), BLASTP, BLASTN, FASTA (Atschul, S.F. et al., J. Molec. Biol. 215:403 (1990), Bestfit program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711 (using the local homology algorithm of Smith and Waterman, Advances in Applied Mathematics 2:482-489 (1981).)

When using any of the sequence alignment programs to determine whether a particular sequence is, for instance, 95% identical to a reference sequence, the parameters are set so that the percentage of identity is calculated over the full length of the reference polynucleotide and that gaps in identity of up to 5% of the total number of nucleotides in the reference polynucleotide are allowed.

A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. 6:237-245 (1990).) The term "sequence" includes nucleotide and amino acid sequences. In a sequence alignment the query and subject sequences are either both nucleotide sequences or both amino acid sequences. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB search of a DNA sequence to calculate percent identity are: Matrix=Unitary, k-tuple=4, Mismatch Penalty=1, Joining Penalty=30, Randomization Group Length=0, and Cutoff Score=1, Gap Penalty=5, Gap Size Penalty 0.05, and Window Size=500 or query sequence length in nucleotide bases, whichever is shorter. Preferred parameters employed to calculate percent identity and similarity of an amino acid alignment are: Matrix=PAM 150, k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty=0.05, and Window Size=500 or query sequence length in amino acid residues, whichever is shorter.

As an illustration, a polynucleotide having a nucleotide sequence of at least 95% "identity" to a sequence contained in SEQ ID NO:X or the cDNA contained in the deposited clone, means that the polynucleotide is identical to a sequence contained in SEQ ID NO:X or the cDNA except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the total length (not just within a given 100 nucleotide stretch). In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to SEQ ID NO:X or the deposited clone, up to 5% of the nucleotides in the sequence contained in SEQ ID NO:X or the cDNA can be deleted, inserted, or substituted with other nucleotides. These changes may occur anywhere throughout the polynucleotide.

Further embodiments of the present invention include polynucleotides having at least 85% identity, more preferably at least 90% identity, and most preferably at least 95%, 96%, 97%, 98% or 99% identity to a sequence contained in SEQ ID NO:X or the cDNA contained in the deposited clone. Of course, due to the degeneracy of the genetic code, one of ordinary skill in the art will immediately recognize that a large number of the polynucleotides having at least 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity



will encode a polypeptide identical to an amino acid sequence contained in SEQ ID NO:Y or the expressed protein produced by the deposited clone.

Similarly, by a polypeptide having an amino acid sequence having at least, for example, 95% "identity" to a reference polypeptide, is intended that the amino acid sequence of the polypeptide is identical to the reference polypeptide except that the polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the total length of the reference polypeptide. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a reference amino acid sequence, up to 5% of the amino acid residues in the reference sequence may be deleted or substituted with another amino acid, or a number of amino acids up to 5% of the total amino acid residues in the reference sequence may be inserted into the reference sequence. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

Further embodiments of the present invention include polypeptides having at least 80% identity, more preferably at least 85% identity, more preferably at least 90% identity, and most preferably at least 95%, 96%, 97%, 98% or 99% identity to an amino acid sequence contained in SEQ ID NO:Y or the expressed protein produced by the deposited clone. Preferably, the above polypeptides should exhibit at least one biological activity of the protein.

In a preferred embodiment, polypeptides of the present invention include polypeptides having at least 90% similarity, more preferably at least 95% similarity, and still more preferably at least 96%, 97%, 98%, or 99% similarity to an amino acid sequence contained in SEQ ID NO:Y or the expressed protein produced by the deposited clone.

The variants may contain alterations in the coding regions, non-coding regions, or both. Especially preferred are polynucleotide variants containing alterations which produce silent substitutions, additions, or deletions, but do not alter the properties or activities of the encoded polypeptide. Nucleotide variants produced by silent substitutions due to the degeneracy of the genetic code are preferred. Moreover, variants in which 5-10, 1-5, or 1-2 amino acids are substituted, deleted, or added in any combination are also preferred. Polynucleotide variants can be produced for a variety of reasons, e.g., to optimize codon expression for a particular host (change codons in the human mRNA to those preferred by a bacterial host such as E. coli).

Naturally occurring variants are called "allelic variants," and refer to one of several alternate forms of a gene occupying a given locus on a chromosome of an

organism. (Genes II, Lewin, B., ed., John Wiley & Sons, New York (1985).) These allelic variants can vary at either the polynucleotide and/or polypeptide level.

Alternatively, non-naturally occurring variants may be produced by mutagenesis techniques or by direct synthesis.

5       Using known methods of protein engineering and recombinant DNA technology, variants may be generated to improve or alter the characteristics of the polypeptides of the present invention. For instance, one or more amino acids can be deleted from the N-terminus or C-terminus of the secreted protein without substantial loss of biological function. The authors of Ron et al., J. Biol. Chem. 268: 2984-2988  
10       (1993), reported variant KGF proteins having heparin binding activity even after deleting 3, 8, or 27 amino-terminal amino acid residues. Similarly, Interferon gamma exhibited up to ten times higher activity after deleting 8-10 amino acid residues from the carboxy terminus of this protein. (Dobeli et al., J. Biotechnology 7:199-216 (1988).)

      Moreover, ample evidence demonstrates that variants often retain a biological  
15       activity similar to that of the naturally occurring protein. For example, Gayle and coworkers (J. Biol. Chem 268:22105-22111 (1993)) conducted extensive mutational analysis of human cytokine IL-1a. They used random mutagenesis to generate over 3,500 individual IL-1a mutants that averaged 2.5 amino acid changes per variant over the entire length of the molecule. Multiple mutations were examined at every possible  
20       amino acid position. The investigators found that "[m]ost of the molecule could be altered with little effect on either [binding or biological activity]." (See, Abstract.) In fact, only 23 unique amino acid sequences, out of more than 3,500 nucleotide sequences examined, produced a protein that significantly differed in activity from wild-type.

25       Furthermore, even if deleting one or more amino acids from the N-terminus or C-terminus of a polypeptide results in modification or loss of one or more biological functions, other biological activities may still be retained. For example, the ability of a deletion variant to induce and/or to bind antibodies which recognize the secreted form will likely be retained when less than the majority of the residues of the secreted form  
30       are removed from the N-terminus or C-terminus. Whether a particular polypeptide lacking N- or C-terminal residues of a protein retains such immunogenic activities can readily be determined by routine methods described herein and otherwise known in the art.

      Thus, the invention further includes polypeptide variants which show  
35       substantial biological activity. Such variants include deletions, insertions, inversions, repeats, and substitutions selected according to general rules known in the art so as have little effect on activity. For example, guidance concerning how to make

phenotypically silent amino acid substitutions is provided in Bowie, J. U. et al., Science 247:1306-1310 (1990), wherein the authors indicate that there are two main strategies for studying the tolerance of an amino acid sequence to change.

The first strategy exploits the tolerance of amino acid substitutions by natural selection during the process of evolution. By comparing amino acid sequences in different species, conserved amino acids can be identified. These conserved amino acids are likely important for protein function. In contrast, the amino acid positions where substitutions have been tolerated by natural selection indicates that these positions are not critical for protein function. Thus, positions tolerating amino acid substitution could be modified while still maintaining biological activity of the protein.

The second strategy uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene to identify regions critical for protein function. For example, site directed mutagenesis or alanine-scanning mutagenesis (introduction of single alanine mutations at every residue in the molecule) can be used. (Cunningham and Wells, Science 244:1081-1085 (1989).) The resulting mutant molecules can then be tested for biological activity.

As the authors state, these two strategies have revealed that proteins are surprisingly tolerant of amino acid substitutions. The authors further indicate which amino acid changes are likely to be permissive at certain amino acid positions in the protein. For example, most buried (within the tertiary structure of the protein) amino acid residues require nonpolar side chains, whereas few features of surface side chains are generally conserved. Moreover, tolerated conservative amino acid substitutions involve replacement of the aliphatic or hydrophobic amino acids Ala, Val, Leu and Ile; replacement of the hydroxyl residues Ser and Thr; replacement of the acidic residues Asp and Glu; replacement of the amide residues Asn and Gln, replacement of the basic residues Lys, Arg, and His; replacement of the aromatic residues Phe, Tyr, and Trp, and replacement of the small-sized amino acids Ala, Ser, Thr, Met, and Gly.

Besides conservative amino acid substitution, variants of the present invention include (i) substitutions with one or more of the non-conserved amino acid residues, where the substituted amino acid residues may or may not be one encoded by the genetic code, or (ii) substitution with one or more of amino acid residues having a substituent group, or (iii) fusion of the mature polypeptide with another compound, such as a compound to increase the stability and/or solubility of the polypeptide (for example, polyethylene glycol), or (iv) fusion of the polypeptide with additional amino acids, such as an IgG Fc fusion region peptide, or leader or secretory sequence, or a sequence facilitating purification. Such variant polypeptides are deemed to be within the scope of those skilled in the art from the teachings herein.

For example, polypeptide variants containing amino acid substitutions of charged amino acids with other charged or neutral amino acids may produce proteins with improved characteristics, such as less aggregation. Aggregation of pharmaceutical formulations both reduces activity and increases clearance due to the aggregate's immunogenic activity. (Pinckard et al., Clin. Exp. Immunol. 2:331-340 (1967); Robbins et al., Diabetes 36: 838-845 (1987); Cleland et al., Crit. Rev. Therapeutic Drug Carrier Systems 10:307-377 (1993).)

### **Polynucleotide and Polypeptide Fragments**

In the present invention, a "polynucleotide fragment" refers to a short polynucleotide having a nucleic acid sequence contained in the deposited clone or shown in SEQ ID NO:X. The short nucleotide fragments are preferably at least about 15 nt, and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably, at least about 40 nt in length. A fragment "at least 20 nt in length," for example, is intended to include 20 or more contiguous bases from the cDNA sequence contained in the deposited clone or the nucleotide sequence shown in SEQ ID NO:X. These nucleotide fragments are useful as diagnostic probes and primers as discussed herein. Of course, larger fragments (e.g., 50, 150, 500, 600, 2000 nucleotides) are preferred.

Moreover, representative examples of polynucleotide fragments of the invention, include, for example, fragments having a sequence from about nucleotide number 1-50, 51-100, 101-150, 151-200, 201-250, 251-300, 301-350, 351-400, 401-450, 451-500, 501-550, 551-600, 651-700, and 701 to the end of SEQ ID NO:X or the cDNA contained in the deposited clone. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) nucleotides, at either terminus or at both termini. Preferably, these fragments encode a polypeptide which has biological activity.

In the present invention, a "polypeptide fragment" refers to a short amino acid sequence contained in SEQ ID NO:Y or encoded by the cDNA contained in the deposited clone. Protein fragments may be "free-standing," or comprised within a larger polypeptide of which the fragment forms a part or region, most preferably as a single continuous region. Representative examples of polypeptide fragments of the invention, include, for example, fragments from about amino acid number 1-20, 21-40, 41-60, 61-80, 81-100, 102-120, 121-140, 141-160, and 161 to the end of the coding region. Moreover, polypeptide fragments can be about 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, or 150 amino acids in length. In this context "about"

includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) amino acids, at either extreme or at both extremes.

Preferred polypeptide fragments include the secreted protein as well as the mature form. Further preferred polypeptide fragments include the secreted protein or the mature form having a continuous series of deleted residues from the amino or the carboxy terminus, or both. For example, any number of amino acids, ranging from 1-60, can be deleted from the amino terminus of either the secreted polypeptide or the mature form. Similarly, any number of amino acids, ranging from 1-30, can be deleted from the carboxy terminus of the secreted protein or mature form. Furthermore, any combination of the above amino and carboxy terminus deletions are preferred. Similarly, polynucleotide fragments encoding these polypeptide fragments are also preferred.

Also preferred are polypeptide and polynucleotide fragments characterized by structural or functional domains, such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions. Polypeptide fragments of SEQ ID NO:Y falling within conserved domains are specifically contemplated by the present invention. Moreover, polynucleotide fragments encoding these domains are also contemplated.

Other preferred fragments are biologically active fragments. Biologically active fragments are those exhibiting activity similar, but not necessarily identical, to an activity of the polypeptide of the present invention. The biological activity of the fragments may include an improved desired activity, or a decreased undesirable activity.

### Epitopes & Antibodies

In the present invention, "epitopes" refer to polypeptide fragments having antigenic or immunogenic activity in an animal, especially in a human. A preferred embodiment of the present invention relates to a polypeptide fragment comprising an epitope, as well as the polynucleotide encoding this fragment. A region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope." In contrast, an "immunogenic epitope" is defined as a part of a protein that elicits an antibody response. (See, for instance, Geysen et al., Proc. Natl. Acad. Sci. USA 81:3998-4002 (1983).)

Fragments which function as epitopes may be produced by any conventional means. (See, e.g., Houghten, R. A., Proc. Natl. Acad. Sci. USA 82:5131-5135 (1985) further described in U.S. Patent No. 4,631,211.)

In the present invention, antigenic epitopes preferably contain a sequence of at least seven, more preferably at least nine, and most preferably between about 15 to about 30 amino acids. Antigenic epitopes are useful to raise antibodies, including monoclonal antibodies, that specifically bind the epitope. (See, for instance, Wilson et al., Cell 37:767-778 (1984); Sutcliffe, J. G. et al., Science 219:660-666 (1983).)

Similarly, immunogenic epitopes can be used to induce antibodies according to methods well known in the art. (See, for instance, Sutcliffe et al., supra; Wilson et al., supra; Chow, M. et al., Proc. Natl. Acad. Sci. USA 82:910-914; and Bittle, F. J. et al., J. Gen. Virol. 66:2347-2354 (1985).) A preferred immunogenic epitope includes the secreted protein. The immunogenic epitopes may be presented together with a carrier protein, such as an albumin, to an animal system (such as rabbit or mouse) or, if it is long enough (at least about 25 amino acids), without a carrier. However, immunogenic epitopes comprising as few as 8 to 10 amino acids have been shown to be sufficient to raise antibodies capable of binding to, at the very least, linear epitopes in a denatured polypeptide (e.g., in Western blotting.)

As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as antibody fragments (such as, for example, Fab and F(ab')<sub>2</sub> fragments) which are capable of specifically binding to protein. Fab and F(ab')<sub>2</sub> fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding than an intact antibody. (Wahl et al., J. Nucl. Med. 24:316-325 (1983).) Thus, these fragments are preferred, as well as the products of a FAB or other immunoglobulin expression library. Moreover, antibodies of the present invention include chimeric, single chain, and humanized antibodies.

### **Fusion Proteins**

Any polypeptide of the present invention can be used to generate fusion proteins. For example, the polypeptide of the present invention, when fused to a second protein, can be used as an antigenic tag. Antibodies raised against the polypeptide of the present invention can be used to indirectly detect the second protein by binding to the polypeptide. Moreover, because secreted proteins target cellular locations based on trafficking signals, the polypeptides of the present invention can be used as targeting molecules once fused to other proteins.

Examples of domains that can be fused to polypeptides of the present invention include not only heterologous signal sequences, but also other heterologous functional regions. The fusion does not necessarily need to be direct, but may occur through linker sequences.

5        Moreover, fusion proteins may also be engineered to improve characteristics of the polypeptide of the present invention. For instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence during purification from the host cell or subsequent handling and storage. Also, peptide moieties may be added to the  
10       polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide. The addition of peptide moieties to facilitate handling of polypeptides are familiar and routine techniques in the art.

      Moreover, polypeptides of the present invention, including fragments, and specifically epitopes, can be combined with parts of the constant domain of  
15       immunoglobulins (IgG), resulting in chimeric polypeptides. These fusion proteins facilitate purification and show an increased half-life in vivo. One reported example describes chimeric proteins consisting of the first two domains of the human CD4-polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins. (EP A 394,827; Traunecker et al., *Nature* 331:84-86  
20       (1988).) Fusion proteins having disulfide-linked dimeric structures (due to the IgG) can also be more efficient in binding and neutralizing other molecules, than the monomeric secreted protein or protein fragment alone. (Fountoulakis et al., *J. Biochem.* 270:3958-3964 (1995).)

      Similarly, EP-A-O 464 533 (Canadian counterpart 2045869) discloses fusion  
25       proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, the Fc part in a fusion protein is beneficial in therapy and diagnosis, and thus can result in, for example, improved pharmacokinetic properties. (EP-A 0232 262.) Alternatively, deleting the Fc part after the fusion protein has been expressed, detected, and purified,  
30       would be desired. For example, the Fc portion may hinder therapy and diagnosis if the fusion protein is used as an antigen for immunizations. In drug discovery, for example, human proteins, such as hIL-5, have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists of hIL-5. (See, D. Bennett et al., *J. Molecular Recognition* 8:52-58 (1995); K. Johanson et al., *J. Biol.*  
35       *Chem.* 270:9459-9471 (1995).)

      Moreover, the polypeptides of the present invention can be fused to marker sequences, such as a peptide which facilitates purification of the fused polypeptide. In

preferred embodiments, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311), among others, many of which are commercially available. As described in Gentz et al., Proc. Natl. Acad. Sci. USA 86:821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein. Another peptide tag useful for purification, the "HA" tag, corresponds to an epitope derived from the influenza hemagglutinin protein. (Wilson et al., Cell 37:767 (1984).)

Thus, any of these above fusions can be engineered using the polynucleotides or the polypeptides of the claimed invention.

### **Vectors, Host Cells, and Protein Production**

The present invention also relates to vectors containing the polynucleotide of the present invention, host cells, and the production of polypeptides by recombinant techniques. The vector may be, for example, a phage, plasmid, viral, or retroviral vector. Retroviral vectors may be replication competent or replication defective. In the latter case, viral propagation generally will occur only in complementing host cells.

The polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector is introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is a virus, it may be packaged in vitro using an appropriate packaging cell line and then transduced into host cells.

The polynucleotide insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the E. coli lac, trp, phoA and tac promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled artisan. The expression constructs will further contain sites for transcription initiation, termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of the transcripts expressed by the constructs will preferably include a translation initiating codon at the beginning and a termination codon (UAA, UGA or UAG) appropriately positioned at the end of the polypeptide to be translated.

As indicated, the expression vectors will preferably include at least one selectable marker. Such markers include dihydrofolate reductase, G418 or neomycin resistance for eukaryotic cell culture and tetracycline, kanamycin or ampicillin resistance genes for culturing in E. coli and other bacteria. Representative examples of appropriate hosts include, but are not limited to, bacterial cells, such as E. coli, Streptomyces and Salmonella typhimurium cells; fungal cells, such as yeast cells; insect cells such as Drosophila S2 and Spodoptera Sf9 cells; animal cells such as CHO, COS,



293, and Bowes melanoma cells; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

Among vectors preferred for use in bacteria include pQE70, pQE60 and pQE-9, available from QIAGEN, Inc.; pBluescript vectors, Phagescript vectors, pNH8A, pNH16a, pNH18A, pNH46A, available from Stratagene Cloning Systems, Inc.; and ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia Biotech, Inc. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Other suitable vectors will be readily apparent to the skilled artisan.

Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, or other methods. Such methods are described in many standard laboratory manuals, such as Davis et al., Basic Methods In Molecular Biology (1986). It is specifically contemplated that the polypeptides of the present invention may in fact be expressed by a host cell lacking a recombinant vector.

A polypeptide of this invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification.

Polypeptides of the present invention, and preferably the secreted form, can also be recovered from: products purified from natural sources, including bodily fluids, tissues and cells, whether directly isolated or cultured; products of chemical synthetic procedures; and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect, and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes. Thus, it is well known in the art that the N-terminal methionine encoded by the translation initiation codon generally is removed with high efficiency from any protein after translation in all eukaryotic cells. While the N-terminal methionine on most proteins also is efficiently removed in most prokaryotes, for some proteins, this prokaryotic removal process is inefficient, depending on the nature of the amino acid to which the N-terminal methionine is covalently linked.

### Uses of the Polynucleotides

Each of the polynucleotides identified herein can be used in numerous ways as reagents. The following description should be considered exemplary and utilizes known techniques.

The polynucleotides of the present invention are useful for chromosome identification. There exists an ongoing need to identify new chromosome markers, since few chromosome marking reagents, based on actual sequence data (repeat polymorphisms), are presently available. Each polynucleotide of the present invention can be used as a chromosome marker.

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the sequences shown in SEQ ID NO:X. Primers can be selected using computer analysis so that primers do not span more than one predicted exon in the genomic DNA. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the SEQ ID NO:X will yield an amplified fragment.

Similarly, somatic hybrids provide a rapid method of PCR mapping the polynucleotides to particular chromosomes. Three or more clones can be assigned per day using a single thermal cycler. Moreover, sublocalization of the polynucleotides can be achieved with panels of specific chromosome fragments. Other gene mapping strategies that can be used include in situ hybridization, prescreening with labeled flow-sorted chromosomes, and preselection by hybridization to construct chromosome specific-cDNA libraries.

Precise chromosomal location of the polynucleotides can also be achieved using fluorescence in situ hybridization (FISH) of a metaphase chromosomal spread. This technique uses polynucleotides as short as 500 or 600 bases; however, polynucleotides 2,000-4,000 bp are preferred. For a review of this technique, see Verma et al., "Human Chromosomes: a Manual of Basic Techniques," Pergamon Press, New York (1988).

For chromosome mapping, the polynucleotides can be used individually (to mark a single chromosome or a single site on that chromosome) or in panels (for marking multiple sites and/or multiple chromosomes). Preferred polynucleotides correspond to the noncoding regions of the cDNAs because the coding sequences are more likely conserved within gene families, thus increasing the chance of cross hybridization during chromosomal mapping.

Once a polynucleotide has been mapped to a precise chromosomal location, the physical position of the polynucleotide can be used in linkage analysis. Linkage

analysis establishes coinheritance between a chromosomal location and presentation of a particular disease. (Disease mapping data are found, for example, in V. McKusick, Mendelian Inheritance in Man (available on line through Johns Hopkins University Welch Medical Library) .) Assuming 1 megabase mapping resolution and one gene per 20 kb, a cDNA precisely localized to a chromosomal region associated with the disease could be one of 50-500 potential causative genes.

Thus, once coinheritance is established, differences in the polynucleotide and the corresponding gene between affected and unaffected individuals can be examined. First, visible structural alterations in the chromosomes, such as deletions or translocations, are examined in chromosome spreads or by PCR. If no structural alterations exist, the presence of point mutations are ascertained. Mutations observed in some or all affected individuals, but not in normal individuals, indicates that the mutation may cause the disease. However, complete sequencing of the polypeptide and the corresponding gene from several normal individuals is required to distinguish the mutation from a polymorphism. If a new polymorphism is identified, this polymorphic polypeptide can be used for further linkage analysis.

Furthermore, increased or decreased expression of the gene in affected individuals as compared to unaffected individuals can be assessed using polynucleotides of the present invention. Any of these alterations (altered expression, chromosomal rearrangement, or mutation) can be used as a diagnostic or prognostic marker.

In addition to the foregoing, a polynucleotide can be used to control gene expression through triple helix formation or antisense DNA or RNA. Both methods rely on binding of the polynucleotide to DNA or RNA. For these techniques, preferred polynucleotides are usually 20 to 40 bases in length and complementary to either the region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 241:456 (1988); and Dervan et al., Science 251:1360 (1991) ) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxy-nucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988).) Triple helix formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques are effective in model systems, and the information disclosed herein can be used to design antisense or triple helix polynucleotides in an effort to treat disease.

Polynucleotides of the present invention are also useful in gene therapy. One goal of gene therapy is to insert a normal gene into an organism having a defective gene, in an effort to correct the genetic defect. The polynucleotides disclosed in the

present invention offer a means of targeting such genetic defects in a highly accurate manner. Another goal is to insert a new gene that was not present in the host genome, thereby producing a new trait in the host cell.

The polynucleotides are also useful for identifying individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identifying personnel. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The polynucleotides of the present invention can be used as additional DNA markers for RFLP.

The polynucleotides of the present invention can also be used as an alternative to RFLP, by determining the actual base-by-base DNA sequence of selected portions of an individual's genome. These sequences can be used to prepare PCR primers for amplifying and isolating such selected DNA, which can then be sequenced. Using this technique, individuals can be identified because each individual will have a unique set of DNA sequences. Once an unique ID database is established for an individual, positive identification of that individual, living or dead, can be made from extremely small tissue samples.

Forensic biology also benefits from using DNA-based identification techniques as disclosed herein. DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, semen, etc., can be amplified using PCR. In one prior art technique, gene sequences amplified from polymorphic loci, such as DQa class II HLA gene, are used in forensic biology to identify individuals. (Erlich, H., PCR Technology, Freeman and Co. (1992).) Once these specific polymorphic loci are amplified, they are digested with one or more restriction enzymes, yielding an identifying set of bands on a Southern blot probed with DNA corresponding to the DQa class II HLA gene. Similarly, polynucleotides of the present invention can be used as polymorphic markers for forensic purposes.

There is also a need for reagents capable of identifying the source of a particular tissue. Such need arises, for example, in forensics when presented with tissue of unknown origin. Appropriate reagents can comprise, for example, DNA probes or primers specific to particular tissue prepared from the sequences of the present invention. Panels of such reagents can identify tissue by species and/or by organ type. In a similar fashion, these reagents can be used to screen tissue cultures for contamination.

In the very least, the polynucleotides of the present invention can be used as molecular weight markers on Southern gels, as diagnostic probes for the presence of a specific mRNA in a particular cell type, as a probe to "subtract-out" known sequences in the process of discovering novel polynucleotides, for selecting and making oligomers for attachment to a "gene chip" or other support, to raise anti-DNA antibodies using DNA immunization techniques, and as an antigen to elicit an immune response.

### Uses of the Polypeptides

Each of the polypeptides identified herein can be used in numerous ways. The following description should be considered exemplary and utilizes known techniques.

A polypeptide of the present invention can be used to assay protein levels in a biological sample using antibody-based techniques. For example, protein expression in tissues can be studied with classical immunohistological methods. (Jalkanen, M., et al., J. Cell. Biol. 101:976-985 (1985); Jalkanen, M., et al., J. Cell. Biol. 105:3087-3096 (1987).) Other antibody-based methods useful for detecting protein gene expression include immunoassays, such as the enzyme linked immunosorbent assay (ELISA) and the radioimmunoassay (RIA). Suitable antibody assay labels are known in the art and include enzyme labels, such as, glucose oxidase, and radioisotopes, such as iodine ( $^{125}\text{I}$ ,  $^{121}\text{I}$ ), carbon ( $^{14}\text{C}$ ), sulfur ( $^{35}\text{S}$ ), tritium ( $^3\text{H}$ ), indium ( $^{112}\text{In}$ ), and technetium ( $^{99\text{m}}\text{Tc}$ ), and fluorescent labels, such as fluorescein and rhodamine, and biotin.

In addition to assaying secreted protein levels in a biological sample, proteins can also be detected in vivo by imaging. Antibody labels or markers for in vivo imaging of protein include those detectable by X-radiography, NMR or ESR. For X-radiography, suitable labels include radioisotopes such as barium or cesium, which emit detectable radiation but are not overtly harmful to the subject. Suitable markers for NMR and ESR include those with a detectable characteristic spin, such as deuterium, which may be incorporated into the antibody by labeling of nutrients for the relevant hybridoma.

A protein-specific antibody or antibody fragment which has been labeled with an appropriate detectable imaging moiety, such as a radioisotope (for example,  $^{131}\text{I}$ ,  $^{112}\text{In}$ ,  $^{99\text{m}}\text{Tc}$ ), a radio-opaque substance, or a material detectable by nuclear magnetic resonance, is introduced (for example, parenterally, subcutaneously, or intraperitoneally) into the mammal. It will be understood in the art that the size of the subject and the imaging system used will determine the quantity of imaging moiety needed to produce diagnostic images. In the case of a radioisotope moiety, for a human subject, the quantity of radioactivity injected will normally range from about 5 to 20

millicuries of  $^{99m}\text{Tc}$ . The labeled antibody or antibody fragment will then preferentially accumulate at the location of cells which contain the specific protein. In vivo tumor imaging is described in S.W. Burchiel et al., "Immunopharmacokinetics of Radiolabeled Antibodies and Their Fragments." (Chapter 13 in Tumor Imaging: The Radiochemical Detection of Cancer, S.W. Burchiel and B. A. Rhodes, eds., Masson Publishing Inc. (1982).)

Thus, the invention provides a diagnostic method of a disorder, which involves (a) assaying the expression of a polypeptide of the present invention in cells or body fluid of an individual; (b) comparing the level of gene expression with a standard gene expression level, whereby an increase or decrease in the assayed polypeptide gene expression level compared to the standard expression level is indicative of a disorder.

Moreover, polypeptides of the present invention can be used to treat disease. For example, patients can be administered a polypeptide of the present invention in an effort to replace absent or decreased levels of the polypeptide (e.g., insulin), to supplement absent or decreased levels of a different polypeptide (e.g., hemoglobin S for hemoglobin B), to inhibit the activity of a polypeptide (e.g., an oncogene), to activate the activity of a polypeptide (e.g., by binding to a receptor), to reduce the activity of a membrane bound receptor by competing with it for free ligand (e.g., soluble TNF receptors used in reducing inflammation), or to bring about a desired response (e.g., blood vessel growth).

Similarly, antibodies directed to a polypeptide of the present invention can also be used to treat disease. For example, administration of an antibody directed to a polypeptide of the present invention can bind and reduce overproduction of the polypeptide. Similarly, administration of an antibody can activate the polypeptide, such as by binding to a polypeptide bound to a membrane (receptor).

At the very least, the polypeptides of the present invention could be used as molecular weight markers on SDS-PAGE gels or on molecular sieve gel filtration columns using methods well known to those of skill in the art. Polypeptides can also be used to raise antibodies, which in turn are used to measure protein expression from a recombinant cell, as a way of assessing transformation of the host cell. Moreover, the polypeptides of the present invention can be used to test the following biological activities.

#### **Biological Activities**

The polynucleotides and polypeptides of the present invention can be used in assays to test for one or more biological activities. If these polynucleotides and polypeptides do exhibit activity in a particular assay, it is likely that these molecules

may be involved in the diseases associated with the biological activity. Thus, the polynucleotides and polypeptides could be used to treat the associated disease.

### Immune Activity

5 A polypeptide or polynucleotide of the present invention may be useful in treating deficiencies or disorders of the immune system, by activating or inhibiting the proliferation, differentiation, or mobilization (chemotaxis) of immune cells. Immune cells develop through a process called hematopoiesis, producing myeloid (platelets, red blood cells, neutrophils, and macrophages) and lymphoid (B and T lymphocytes) cells  
10 from pluripotent stem cells. The etiology of these immune deficiencies or disorders may be genetic, somatic, such as cancer or some autoimmune disorders, acquired (e.g., by chemotherapy or toxins), or infectious. Moreover, a polynucleotide or polypeptide of the present invention can be used as a marker or detector of a particular immune system disease or disorder.

15 A polynucleotide or polypeptide of the present invention may be useful in treating or detecting deficiencies or disorders of hematopoietic cells. A polypeptide or polynucleotide of the present invention could be used to increase differentiation and proliferation of hematopoietic cells, including the pluripotent stem cells, in an effort to treat those disorders associated with a decrease in certain (or many) types hematopoietic  
20 cells. Examples of immunologic deficiency syndromes include, but are not limited to: blood protein disorders (e.g. agammaglobulinemia, dysgammaglobulinemia), ataxia telangiectasia, common variable immunodeficiency, Digeorge Syndrome, HIV infection, HTLV-BLV infection, leukocyte adhesion deficiency syndrome, lymphopenia, phagocyte bactericidal dysfunction, severe combined immunodeficiency (SCIDs), Wiskott-Aldrich Disorder, anemia, thrombocytopenia, or hemoglobinuria.

25 Moreover, a polypeptide or polynucleotide of the present invention could also be used to modulate hemostatic (the stopping of bleeding) or thrombolytic activity (clot formation). For example, by increasing hemostatic or thrombolytic activity, a polynucleotide or polypeptide of the present invention could be used to treat blood  
30 coagulation disorders (e.g., afibrinogenemia, factor deficiencies), blood platelet disorders (e.g. thrombocytopenia), or wounds resulting from trauma, surgery, or other causes. Alternatively, a polynucleotide or polypeptide of the present invention that can decrease hemostatic or thrombolytic activity could be used to inhibit or dissolve clotting. These molecules could be important in the treatment of heart attacks  
35 (infarction), strokes, or scarring.

A polynucleotide or polypeptide of the present invention may also be useful in treating or detecting autoimmune disorders. Many autoimmune disorders result from

inappropriate recognition of self as foreign material by immune cells. This inappropriate recognition results in an immune response leading to the destruction of the host tissue. Therefore, the administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing autoimmune disorders.

Examples of autoimmune disorders that can be treated or detected by the present invention include, but are not limited to: Addison's Disease, hemolytic anemia, antiphospholipid syndrome, rheumatoid arthritis, dermatitis, allergic encephalomyelitis, glomerulonephritis, Goodpasture's Syndrome, Graves' Disease, Multiple Sclerosis, Myasthenia Gravis, Neuritis, Ophthalmia, Bullous Pemphigoid, Pemphigus, Polyendocrinopathies, Purpura, Reiter's Disease, Stiff-Man Syndrome, Autoimmune Thyroiditis, Systemic Lupus Erythematosus, Autoimmune Pulmonary Inflammation, Guillain-Barre Syndrome, insulin dependent diabetes mellitus, and autoimmune inflammatory eye disease.

Similarly, allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems, may also be treated by a polypeptide or polynucleotide of the present invention. Moreover, these molecules can be used to treat anaphylaxis, hypersensitivity to an antigenic molecule, or blood group incompatibility.

A polynucleotide or polypeptide of the present invention may also be used to treat and/or prevent organ rejection or graft-versus-host disease (GVHD). Organ rejection occurs by host immune cell destruction of the transplanted tissue through an immune response. Similarly, an immune response is also involved in GVHD, but, in this case, the foreign transplanted immune cells destroy the host tissues. The administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing organ rejection or GVHD.

Similarly, a polypeptide or polynucleotide of the present invention may also be used to modulate inflammation. For example, the polypeptide or polynucleotide may inhibit the proliferation and differentiation of cells involved in an inflammatory response. These molecules can be used to treat inflammatory conditions, both chronic and acute conditions, including inflammation associated with infection (e.g., septic shock, sepsis, or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine induced lung injury, inflammatory bowel disease, Crohn's disease, or resulting from over production of cytokines (e.g., TNF or IL-1.)



### **Hyperproliferative Disorders**

A polypeptide or polynucleotide can be used to treat or detect hyperproliferative disorders, including neoplasms. A polypeptide or polynucleotide of the present invention may inhibit the proliferation of the disorder through direct or indirect interactions. Alternatively, a polypeptide or polynucleotide of the present invention may proliferate other cells which can inhibit the hyperproliferative disorder.

For example, by increasing an immune response, particularly increasing antigenic qualities of the hyperproliferative disorder or by proliferating, differentiating, or mobilizing T-cells, hyperproliferative disorders can be treated. This immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, decreasing an immune response may also be a method of treating hyperproliferative disorders, such as a chemotherapeutic agent.

Examples of hyperproliferative disorders that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but are not limited to neoplasms located in the: abdomen, bone, breast, digestive system, liver, pancreas, peritoneum, endocrine glands (adrenal, parathyroid, pituitary, testicles, ovary, thymus, thyroid), eye, head and neck, nervous (central and peripheral), lymphatic system, pelvic, skin, soft tissue, spleen, thoracic, and urogenital.

Similarly, other hyperproliferative disorders can also be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of such hyperproliferative disorders include, but are not limited to: hypergammaglobulinemia, lymphoproliferative disorders, paraproteinemias, purpura, sarcoidosis, Sezary Syndrome, Waldenstrom's Macroglobulinemia, Gaucher's Disease, histiocytosis, and any other hyperproliferative disease, besides neoplasia, located in an organ system listed above.

### **Infectious Disease**

A polypeptide or polynucleotide of the present invention can be used to treat or detect infectious agents. For example, by increasing the immune response, particularly increasing the proliferation and differentiation of B and/or T cells, infectious diseases may be treated. The immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, the polypeptide or polynucleotide of the present invention may also directly inhibit the infectious agent, without necessarily eliciting an immune response.

Viruses are one example of an infectious agent that can cause disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of viruses, include, but are not limited to the following DNA and RNA viral families: Arbovirus, Adenoviridae, Arenaviridae, Arterivirus, B

5 Birnaviridae, Bunyaviridae, Caliciviridae, Circoviridae, Coronaviridae, Flaviviridae, Hepadnaviridae (Hepatitis), Herpesviridae (such as, Cytomegalovirus, Herpes Simplex, Herpes Zoster), Mononegavirus (e.g., Paramyxoviridae, Morbillivirus, Rhabdoviridae), Orthomyxoviridae (e.g., Influenza), Papovaviridae, Parvoviridae, Picornaviridae, Poxviridae (such as Smallpox or Vaccinia), Reoviridae (e.g.,

10 Rotavirus), Retroviridae (HTLV-I, HTLV-II, Lentivirus), and Togaviridae (e.g., Rubivirus). Viruses falling within these families can cause a variety of diseases or symptoms, including, but not limited to: arthritis, bronchiolitis, encephalitis, eye infections (e.g., conjunctivitis, keratitis), chronic fatigue syndrome, hepatitis (A, B, C, E, Chronic Active, Delta), meningitis, opportunistic infections (e.g., AIDS),

15 pneumonia, Burkitt's Lymphoma, chickenpox, hemorrhagic fever, Measles, Mumps, Parainfluenza, Rabies, the common cold, Polio, leukemia, Rubella, sexually transmitted diseases, skin diseases (e.g., Kaposi's, warts), and viremia. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

20 Similarly, bacterial or fungal agents that can cause disease or symptoms and that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following Gram-Negative and Gram-positive bacterial families and fungi: Actinomycetales (e.g., Corynebacterium, Mycobacterium, Norcardia), Aspergillosis, Bacillaceae (e.g., Anthrax, Clostridium), Bacteroidaceae,

25 Blastomycosis, Bordetella, Borrelia, Brucellosis, Candidiasis, Campylobacter, Coccidioidomycosis, Cryptococcosis, Dermatocycoses, Enterobacteriaceae (Klebsiella, Salmonella, Serratia, Yersinia), Erysipelothrix, Helicobacter, Legionellosis, Leptospirosis, Listeria, Mycoplasmatales, Neisseriaceae (e.g., Acinetobacter, Gonorrhea, Meningococcal), Pasteurellaceae Infections (e.g., Actinobacillus,

30 Heamophilus, Pasteurella), Pseudomonas, Rickettsiaceae, Chlamydiaceae, Syphilis, and Staphylococcal. These bacterial or fungal families can cause the following diseases or symptoms, including, but not limited to: bacteremia, endocarditis, eye infections (conjunctivitis, tuberculosis, uveitis), gingivitis, opportunistic infections (e.g., AIDS related infections), paronychia, prosthesis-related infections, Reiter's Disease,

35 respiratory tract infections, such as Whooping Cough or Empyema, sepsis, Lyme Disease, Cat-Scratch Disease, Dysentery, Paratyphoid Fever, food poisoning, Typhoid, pneumonia, Gonorrhea, meningitis, Chlamydia, Syphilis, Diphtheria,

Leprosy, Paratuberculosis, Tuberculosis, Lupus, Botulism, gangrene, tetanus, impetigo, Rheumatic Fever, Scarlet Fever, sexually transmitted diseases, skin diseases (e.g., cellulitis, dermatocycoses), toxemia, urinary tract infections, wound infections. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Moreover, parasitic agents causing disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following families: Amebiasis, Babesiosis, Coccidiosis, Cryptosporidiosis, Dientamoebiasis, Dourine, Ectoparasitic, Giardiasis, Helminthiasis, Leishmaniasis, Theileriasis, Toxoplasmosis, Trypanosomiasis, and Trichomonas. These parasites can cause a variety of diseases or symptoms, including, but not limited to: Scabies, Trombiculiasis, eye infections, intestinal disease (e.g., dysentery, giardiasis), liver disease, lung disease, opportunistic infections (e.g., AIDS related), Malaria, pregnancy complications, and toxoplasmosis. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Preferably, treatment using a polypeptide or polynucleotide of the present invention could either be by administering an effective amount of a polypeptide to the patient, or by removing cells from the patient, supplying the cells with a polynucleotide of the present invention, and returning the engineered cells to the patient (ex vivo therapy). Moreover, the polypeptide or polynucleotide of the present invention can be used as an antigen in a vaccine to raise an immune response against infectious disease.

### **Regeneration**

A polynucleotide or polypeptide of the present invention can be used to differentiate, proliferate, and attract cells, leading to the regeneration of tissues. (See, Science 276:59-87 (1997).) The regeneration of tissues could be used to repair, replace, or protect tissue damaged by congenital defects, trauma (wounds, burns, incisions, or ulcers), age, disease (e.g. osteoporosis, osteoarthritis, periodontal disease, liver failure), surgery, including cosmetic plastic surgery, fibrosis, reperfusion injury, or systemic cytokine damage.

Tissues that could be regenerated using the present invention include organs (e.g., pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac), vascular (including vascular endothelium), nervous, hematopoietic, and skeletal (bone, cartilage, tendon, and ligament) tissue. Preferably, regeneration occurs without or decreased scarring. Regeneration also may include angiogenesis.

Moreover, a polynucleotide or polypeptide of the present invention may increase regeneration of tissues difficult to heal. For example, increased tendon/ligament regeneration would quicken recovery time after damage. A polynucleotide or polypeptide of the present invention could also be used prophylactically in an effort to avoid damage. Specific diseases that could be treated include of tendinitis, carpal tunnel syndrome, and other tendon or ligament defects. A further example of tissue regeneration of non-healing wounds includes pressure ulcers, ulcers associated with vascular insufficiency, surgical, and traumatic wounds.

Similarly, nerve and brain tissue could also be regenerated by using a polynucleotide or polypeptide of the present invention to proliferate and differentiate nerve cells. Diseases that could be treated using this method include central and peripheral nervous system diseases, neuropathies, or mechanical and traumatic disorders (e.g., spinal cord disorders, head trauma, cerebrovascular disease, and stroke). Specifically, diseases associated with peripheral nerve injuries, peripheral neuropathy (e.g., resulting from chemotherapy or other medical therapies), localized neuropathies, and central nervous system diseases (e.g., Alzheimer's disease, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome), could all be treated using the polynucleotide or polypeptide of the present invention.

### Chemotaxis

A polynucleotide or polypeptide of the present invention may have chemotaxis activity. A chemotactic molecule attracts or mobilizes cells (e.g., monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells) to a particular site in the body, such as inflammation, infection, or site of hyperproliferation. The mobilized cells can then fight off and/or heal the particular trauma or abnormality.

A polynucleotide or polypeptide of the present invention may increase chemotactic activity of particular cells. These chemotactic molecules can then be used to treat inflammation, infection, hyperproliferative disorders, or any immune system disorder by increasing the number of cells targeted to a particular location in the body. For example, chemotactic molecules can be used to treat wounds and other trauma to tissues by attracting immune cells to the injured location. Chemotactic molecules of the present invention can also attract fibroblasts, which can be used to treat wounds.

It is also contemplated that a polynucleotide or polypeptide of the present invention may inhibit chemotactic activity. These molecules could also be used to treat

disorders. Thus, a polynucleotide or polypeptide of the present invention could be used as an inhibitor of chemotaxis.

### **Binding Activity**

5 A polypeptide of the present invention may be used to screen for molecules that bind to the polypeptide or for molecules to which the polypeptide binds. The binding of the polypeptide and the molecule may activate (agonist), increase, inhibit (antagonist), or decrease activity of the polypeptide or the molecule bound. Examples of such molecules include antibodies, oligonucleotides, proteins (e.g., receptors), or  
10 small molecules.

Preferably, the molecule is closely related to the natural ligand of the polypeptide, e.g., a fragment of the ligand, or a natural substrate, a ligand, a structural or functional mimetic. (See, Coligan et al., Current Protocols in Immunology 1(2):Chapter 5 (1991).) Similarly, the molecule can be closely related to the natural  
15 receptor to which the polypeptide binds, or at least, a fragment of the receptor capable of being bound by the polypeptide (e.g., active site). In either case, the molecule can be rationally designed using known techniques.

Preferably, the screening for these molecules involves producing appropriate cells which express the polypeptide, either as a secreted protein or on the cell  
20 membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing the polypeptide (or cell membrane containing the expressed polypeptide) are then preferably contacted with a test compound potentially containing the molecule to observe binding, stimulation, or inhibition of activity of either the polypeptide or the molecule.

25 The assay may simply test binding of a candidate compound to the polypeptide, wherein binding is detected by a label, or in an assay involving competition with a labeled competitor. Further, the assay may test whether the candidate compound results in a signal generated by binding to the polypeptide.

Alternatively, the assay can be carried out using cell-free preparations,  
30 polypeptide/molecule affixed to a solid support, chemical libraries, or natural product mixtures. The assay may also simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide, measuring polypeptide/molecule activity or binding, and comparing the polypeptide/molecule activity or binding to a standard.

35 Preferably, an ELISA assay can measure polypeptide level or activity in a sample (e.g., biological sample) using a monoclonal or polyclonal antibody. The

antibody can measure polypeptide level or activity by either binding, directly or indirectly, to the polypeptide or by competing with the polypeptide for a substrate.

All of these above assays can be used as diagnostic or prognostic markers. The molecules discovered using these assays can be used to treat disease or to bring about a particular result in a patient (e.g., blood vessel growth) by activating or inhibiting the polypeptide/molecule. Moreover, the assays can discover agents which may inhibit or enhance the production of the polypeptide from suitably manipulated cells or tissues.

Therefore, the invention includes a method of identifying compounds which bind to a polypeptide of the invention comprising the steps of: (a) incubating a candidate binding compound with a polypeptide of the invention; and (b) determining if binding has occurred. Moreover, the invention includes a method of identifying agonists/antagonists comprising the steps of: (a) incubating a candidate compound with a polypeptide of the invention, (b) assaying a biological activity, and (b) determining if a biological activity of the polypeptide has been altered.

#### **Other Activities**

A polypeptide or polynucleotide of the present invention may also increase or decrease the differentiation or proliferation of embryonic stem cells, besides, as discussed above, hematopoietic lineage.

A polypeptide or polynucleotide of the present invention may also be used to modulate mammalian characteristics, such as body height, weight, hair color, eye color, skin, percentage of adipose tissue, pigmentation, size, and shape (e.g., cosmetic surgery). Similarly, a polypeptide or polynucleotide of the present invention may be used to modulate mammalian metabolism affecting catabolism, anabolism, processing, utilization, and storage of energy.

A polypeptide or polynucleotide of the present invention may be used to change a mammal's mental state or physical state by influencing biorhythms, circadian rhythms, depression (including depressive disorders), tendency for violence, tolerance for pain, reproductive capabilities (preferably by Activin or Inhibin-like activity), hormonal or endocrine levels, appetite, libido, memory, stress, or other cognitive qualities.

A polypeptide or polynucleotide of the present invention may also be used as a food additive or preservative, such as to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional components.

**Other Preferred Embodiments**

Other preferred embodiments of the claimed invention include an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 50 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Clone Sequence and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Start Codon and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Similarly preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 150 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

Further preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 500 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

A further preferred embodiment is a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the nucleotide sequence of SEQ ID NO:X beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence of SEQ ID NO:X.

Also preferred is an isolated nucleic acid molecule which hybridizes under stringent hybridization conditions to a nucleic acid molecule, wherein said nucleic acid molecule which hybridizes does not hybridize under stringent hybridization conditions to a nucleic acid molecule having a nucleotide sequence consisting of only A residues or of only T residues.

Also preferred is a composition of matter comprising a DNA molecule which comprises a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the material deposited with the American Type Culture Collection and given the ATCC Deposit Number shown in Table 1 for said cDNA Clone Identifier.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in the nucleotide sequence of a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the deposit given the ATCC Deposit Number shown in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said sequence of at least 50 contiguous nucleotides is included in the nucleotide sequence of the complete open reading frame sequence encoded by said human cDNA clone.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 150 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 500 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is a method for detecting in a biological sample a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method comprises a step of comparing a nucleotide sequence of at least one nucleic acid molecule in said sample with a sequence selected from said group and determining



whether the sequence of said nucleic acid molecule in said sample is at least 95% identical to said selected sequence.

Also preferred is the above method wherein said step of comparing sequences comprises determining the extent of nucleic acid hybridization between nucleic acid molecules in said sample and a nucleic acid molecule comprising said sequence selected from said group. Similarly, also preferred is the above method wherein said step of comparing sequences is performed by comparing the nucleotide sequence determined from a nucleic acid molecule in said sample with said sequence selected from said group. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

A further preferred embodiment is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting nucleic acid molecules in said sample, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

The method for identifying the species, tissue or cell type of a biological sample can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject nucleic acid molecules, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

The method for diagnosing a pathological condition can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a composition of matter comprising isolated nucleic acid molecules wherein the nucleotide sequences of said nucleic acid molecules comprise a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1.

Also preferred is a polypeptide, wherein said sequence of contiguous amino acids is included in the amino acid sequence of SEQ ID NO:Y in the range of positions beginning with the residue at about the position of the First Amino Acid of the Secreted Portion and ending with the residue at about the Last Amino Acid of the Open Reading Frame as set forth for SEQ ID NO:Y in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the complete amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is a polypeptide wherein said sequence of contiguous amino acids is included in the amino acid sequence of a secreted portion of the secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the

amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at  
5 least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at  
10 least 95% identical to the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is an isolated antibody which binds specifically to a  
15 polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in  
20 the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method for detecting in a biological sample a polypeptide comprising an amino acid sequence which is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1;  
25 and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method comprises a step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group and determining  
30 whether the sequence of said polypeptide molecule in said sample is at least 90% identical to said sequence of at least 10 contiguous amino acids.

Also preferred is the above method wherein said step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group comprises determining the extent of specific binding of  
35 polypeptides in said sample to an antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an

amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

5 Also preferred is the above method wherein said step of comparing sequences is performed by comparing the amino acid sequence determined from a polypeptide molecule in said sample with said sequence selected from said group.

Also preferred is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting polypeptide molecules in said sample, if any, comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method for identifying the species, tissue or cell type of a biological sample, which method comprises a step of detecting polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the above group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

In any of these methods, the step of detecting said polypeptide molecules includes using an antibody.

35 Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a nucleotide sequence encoding a polypeptide wherein said polypeptide comprises an amino acid sequence that is at least

90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said nucleotide sequence encoding a polypeptide has been optimized for expression of said polypeptide in a prokaryotic host.

Also preferred is an isolated nucleic acid molecule, wherein said polypeptide comprises an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method of making a recombinant vector comprising inserting any of the above isolated nucleic acid molecule into a vector. Also preferred is the recombinant vector produced by this method. Also preferred is a method of making a recombinant host cell comprising introducing the vector into a host cell, as well as the recombinant host cell produced by this method.

Also preferred is a method of making an isolated polypeptide comprising culturing this recombinant host cell under conditions such that said polypeptide is expressed and recovering said polypeptide. Also preferred is this method of making an isolated polypeptide, wherein said recombinant host cell is a eukaryotic cell and said polypeptide is a secreted portion of a human secreted protein comprising an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y beginning with the residue at the position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y wherein Y is an integer set forth in Table 1 and said position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y is defined in Table 1; and an amino acid sequence of a secreted portion of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The isolated polypeptide produced by this method is also preferred.

Also preferred is a method of treatment of an individual in need of an increased level of a secreted protein activity, which method comprises administering to such an individual a pharmaceutical composition comprising an amount of an isolated

polypeptide, polynucleotide, or antibody of the claimed invention effective to increase the level of said protein activity in said individual.

Having generally described the invention, the same will be more readily understood by reference to the following examples, which are provided by way of illustration and are not intended as limiting.

### Examples

#### Example 1: Isolation of a Selected cDNA Clone From the Deposited Sample

Each cDNA clone in a cited ATCC deposit is contained in a plasmid vector. Table 1 identifies the vectors used to construct the cDNA library from which each clone was isolated. In many cases, the vector used to construct the library is a phage vector from which a plasmid has been excised. The table immediately below correlates the related plasmid for each phage vector used in constructing the cDNA library. For example, where a particular clone is identified in Table 1 as being isolated in the vector "Lambda Zap," the corresponding deposited clone is in "pBluescript."

	<u>Vector Used to Construct Library</u>	<u>Corresponding Deposited Plasmid</u>
	Lambda Zap	pBluescript (pBS)
	Uni-Zap XR	pBluescript (pBS)
20	Zap Express	pBK
	lafmid BA	plafmid BA
	pSport1	pSport1
	pCMVSPORT 2.0	pCMVSPORT 2.0
	pCMVSPORT 3.0	pCMVSPORT 3.0
25	pCR®2.1	pCR®2.1

Vectors Lambda Zap (U.S. Patent Nos. 5,128,256 and 5,286,636), Uni-Zap XR (U.S. Patent Nos. 5,128, 256 and 5,286,636), Zap Express (U.S. Patent Nos. 5,128,256 and 5,286,636), pBluescript (pBS) (Short, J. M. et al., Nucleic Acids Res. 16:7583-7600 (1988); Alting-Mees, M. A. and Short, J. M., Nucleic Acids Res. 17:9494 (1989)) and pBK (Alting-Mees, M. A. et al., Strategies 5:58-61 (1992)) are commercially available from Stratagene Cloning Systems, Inc., 11011 N. Torrey Pines Road, La Jolla, CA, 92037. pBS contains an ampicillin resistance gene and pBK contains a neomycin resistance gene. Both can be transformed into E. coli strain XL-1 Blue, also available from Stratagene. pBS comes in 4 forms SK+, SK-, KS+ and KS-. The S and K refers to the orientation of the polylinker to the T7 and T3 primer sequences which flank the polylinker region ("S" is for SacI and "K" is for KpnI which are the first sites on each respective end of the linker). "+" or "-" refer to the orientation

of the f1 origin of replication ("ori"), such that in one orientation, single stranded rescue initiated from the f1 ori generates sense strand DNA and in the other, antisense.

Vectors pSport1, pCMVSPORT 2.0 and pCMVSPORT 3.0, were obtained from Life Technologies, Inc., P. O. Box 6009, Gaithersburg, MD 20897. All Sport vectors  
5 contain an ampicillin resistance gene and may be transformed into E. coli strain DH10B, also available from Life Technologies. (See, for instance, Gruber, C. E., et al., Focus 15:59 (1993).) Vector lafmid BA (Bento Soares, Columbia University, NY) contains an ampicillin resistance gene and can be transformed into E. coli strain XL-1 Blue. Vector pCR<sup>®</sup>2.1, which is available from Invitrogen, 1600 Faraday Avenue,  
10 Carlsbad, CA 92008, contains an ampicillin resistance gene and may be transformed into E. coli strain DH10B, available from Life Technologies. (See, for instance, Clark, J. M., Nuc. Acids Res. 16:9677-9686 (1988) and Mead, D. et al., Bio/Technology 9: (1991).) Preferably, a polynucleotide of the present invention does not comprise the phage vector sequences identified for the particular clone in Table 1, as well as the  
15 corresponding plasmid vector sequences designated above.

The deposited material in the sample assigned the ATCC Deposit Number cited in Table 1 for any given cDNA clone also may contain one or more additional plasmids, each comprising a cDNA clone different from that given clone. Thus, deposits sharing the same ATCC Deposit Number contain at least a plasmid for each cDNA clone  
20 identified in Table 1. Typically, each ATCC deposit sample cited in Table 1 comprises a mixture of approximately equal amounts (by weight) of about 50 plasmid DNAs, each containing a different cDNA clone; but such a deposit sample may include plasmids for more or less than 50 cDNA clones, up to about 500 cDNA clones.

Two approaches can be used to isolate a particular clone from the deposited  
25 sample of plasmid DNAs cited for that clone in Table 1. First, a plasmid is directly isolated by screening the clones using a polynucleotide probe corresponding to SEQ ID NO:X.

Particularly, a specific polynucleotide with 30-40 nucleotides is synthesized using an Applied Biosystems DNA synthesizer according to the sequence reported.  
30 The oligonucleotide is labeled, for instance, with <sup>32</sup>P-γ-ATP using T4 polynucleotide kinase and purified according to routine methods. (E.g., Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring, NY (1982).) The plasmid mixture is transformed into a suitable host, as indicated above (such as XL-1 Blue (Stratagene)) using techniques known to those of skill in the art, such as  
35 those provided by the vector supplier or in related publications or patents cited above. The transformants are plated on 1.5% agar plates (containing the appropriate selection

agent, e.g., ampicillin) to a density of about 150 transformants (colonies) per plate. These plates are screened using Nylon membranes according to routine methods for bacterial colony screening (e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd Edit., (1989), Cold Spring Harbor Laboratory Press, pages 1.93 to 1.104), or other techniques known to those of skill in the art.

Alternatively, two primers of 17-20 nucleotides derived from both ends of the SEQ ID NO:X (i.e., within the region of SEQ ID NO:X bounded by the 5' NT and the 3' NT of the clone defined in Table 1) are synthesized and used to amplify the desired cDNA using the deposited cDNA plasmid as a template. The polymerase chain reaction is carried out under routine conditions, for instance, in 25  $\mu$ l of reaction mixture with 0.5  $\mu$ g of the above cDNA template. A convenient reaction mixture is 1.5-5 mM  $MgCl_2$ , 0.01% (w/v) gelatin, 20  $\mu$ M each of dATP, dCTP, dGTP, dTTP, 25 pmol of each primer and 0.25 Unit of Taq polymerase. Thirty five cycles of PCR (denaturation at 94°C for 1 min; annealing at 55°C for 1 min; elongation at 72°C for 1 min) are performed with a Perkin-Elmer Cetus automated thermal cycler. The amplified product is analyzed by agarose gel electrophoresis and the DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the selected sequence by subcloning and sequencing the DNA product.

Several methods are available for the identification of the 5' or 3' non-coding portions of a gene which may not be present in the deposited clone. These methods include but are not limited to, filter probing, clone enrichment using specific probes, and protocols similar or identical to 5' and 3' "RACE" protocols which are well known in the art. For instance, a method similar to 5' RACE is available for generating the missing 5' end of a desired full-length transcript. (Fromont-Racine et al., *Nucleic Acids Res.* 21(7):1683-1684 (1993).)

Briefly, a specific RNA oligonucleotide is ligated to the 5' ends of a population of RNA presumably containing full-length gene RNA transcripts. A primer set containing a primer specific to the ligated RNA oligonucleotide and a primer specific to a known sequence of the gene of interest is used to PCR amplify the 5' portion of the desired full-length gene. This amplified product may then be sequenced and used to generate the full length gene.

This above method starts with total RNA isolated from the desired source, although poly-A+ RNA can be used. The RNA preparation can then be treated with phosphatase if necessary to eliminate 5' phosphate groups on degraded or damaged RNA which may interfere with the later RNA ligase step. The phosphatase should then be inactivated and the RNA treated with tobacco acid pyrophosphatase in order to



remove the cap structure present at the 5' ends of messenger RNAs. This reaction leaves a 5' phosphate group at the 5' end of the cap cleaved RNA which can then be ligated to an RNA oligonucleotide using T4 RNA ligase.

5 This modified RNA preparation is used as a template for first strand cDNA synthesis using a gene specific oligonucleotide. The first strand synthesis reaction is used as a template for PCR amplification of the desired 5' end using a primer specific to the ligated RNA oligonucleotide and a primer specific to the known sequence of the gene of interest. The resultant product is then sequenced and analyzed to confirm that the 5' end sequence belongs to the desired gene.

10

**Example 2: Isolation of Genomic Clones Corresponding to a Polynucleotide**

A human genomic P1 library (Genomic Systems, Inc.) is screened by PCR using primers selected for the cDNA sequence corresponding to SEQ ID NO:X.,  
15 according to the method described in Example 1. (See also, Sambrook.)

**Example 3: Tissue Distribution of Polypeptide**

Tissue distribution of mRNA expression of polynucleotides of the present invention is determined using protocols for Northern blot analysis, described by,  
20 among others, Sambrook et al. For example, a cDNA probe produced by the method described in Example 1 is labeled with P<sup>32</sup> using the rediprime™ DNA labeling system (Amersham Life Science), according to manufacturer's instructions. After labeling, the probe is purified using CHROMA SPIN-100™ column (Clontech Laboratories, Inc.), according to manufacturer's protocol number PT1200-1. The purified labeled probe is  
25 then used to examine various human tissues for mRNA expression.

Multiple Tissue Northern (MTN) blots containing various human tissues (H) or human immune system tissues (IM) (Clontech) are examined with the labeled probe using ExpressHyb™ hybridization solution (Clontech) according to manufacturer's protocol number PT1190-1. Following hybridization and washing, the blots are  
30 mounted and exposed to film at -70°C overnight, and the films developed according to standard procedures.

**Example 4: Chromosomal Mapping of the Polynucleotides**

An oligonucleotide primer set is designed according to the sequence at the 5'  
35 end of SEQ ID NO:X. This primer preferably spans about 100 nucleotides. This primer set is then used in a polymerase chain reaction under the following set of

conditions : 30 seconds, 95°C; 1 minute, 56°C; 1 minute, 70°C. This cycle is repeated 32 times followed by one 5 minute cycle at 70°C. Human, mouse, and hamster DNA is used as template in addition to a somatic cell hybrid panel containing individual chromosomes or chromosome fragments (Bios, Inc). The reactions is analyzed on  
5 either 8% polyacrylamide gels or 3.5 % agarose gels. Chromosome mapping is determined by the presence of an approximately 100 bp PCR fragment in the particular somatic cell hybrid.

#### **Example 5: Bacterial Expression of a Polypeptide**

10 A polynucleotide encoding a polypeptide of the present invention is amplified using PCR oligonucleotide primers corresponding to the 5' and 3' ends of the DNA sequence, as outlined in Example 1, to synthesize insertion fragments. The primers used to amplify the cDNA insert should preferably contain restriction sites, such as BamHI and XbaI, at the 5' end of the primers in order to clone the amplified product  
15 into the expression vector. For example, BamHI and XbaI correspond to the restriction enzyme sites on the bacterial expression vector pQE-9. (Qiagen, Inc., Chatsworth, CA). This plasmid vector encodes antibiotic resistance (Amp<sup>r</sup>), a bacterial origin of replication (ori), an IPTG-regulatable promoter/operator (P/O), a ribosome binding site (RBS), a 6-histidine tag (6-His), and restriction enzyme cloning sites.

20 The pQE-9 vector is digested with BamHI and XbaI and the amplified fragment is ligated into the pQE-9 vector maintaining the reading frame initiated at the bacterial RBS. The ligation mixture is then used to transform the E. coli strain M15/rep4 (Qiagen, Inc.) which contains multiple copies of the plasmid pREP4, which expresses the lacI repressor and also confers kanamycin resistance (Kan<sup>r</sup>). Transformants are  
25 identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies are selected. Plasmid DNA is isolated and confirmed by restriction analysis.

Clones containing the desired constructs are grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 ug/ml) and Kan (25 ug/ml). The O/N culture is used to inoculate a large culture at a ratio of 1:100 to 1:250. The  
30 cells are grown to an optical density 600 (O.D.<sup>600</sup>) of between 0.4 and 0.6. IPTG (Isopropyl-B-D-thiogalacto pyranoside) is then added to a final concentration of 1 mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene expression.

Cells are grown for an extra 3 to 4 hours. Cells are then harvested by  
35 centrifugation (20 mins at 6000Xg). The cell pellet is solubilized in the chaotropic

agent 6 Molar Guanidine HCl by stirring for 3-4 hours at 4°C. The cell debris is removed by centrifugation, and the supernatant containing the polypeptide is loaded onto a nickel-nitrilo-tri-acetic acid ("Ni-NTA") affinity resin column (available from QIAGEN, Inc., *supra*). Proteins with a 6 x His tag bind to the Ni-NTA resin with high affinity and can be purified in a simple one-step procedure (for details see: The

5 QIAexpressionist (1995) QIAGEN, Inc., *supra*).  
Briefly, the supernatant is loaded onto the column in 6 M guanidine-HCl, pH 8, the column is first washed with 10 volumes of 6 M guanidine-HCl, pH 8, then washed with 10 volumes of 6 M guanidine-HCl pH 6, and finally the polypeptide is eluted with  
10 6 M guanidine-HCl, pH 5.

The purified protein is then renatured by dialyzing it against phosphate-buffered saline (PBS) or 50 mM Na-acetate, pH 6 buffer plus 200 mM NaCl. Alternatively, the protein can be successfully refolded while immobilized on the Ni-NTA column. The recommended conditions are as follows: renature using a linear 6M-1M urea gradient in  
15 500 mM NaCl, 20% glycerol, 20 mM Tris/HCl pH 7.4, containing protease inhibitors. The renaturation should be performed over a period of 1.5 hours or more. After renaturation the proteins are eluted by the addition of 250 mM imidazole. Imidazole is removed by a final dialyzing step against PBS or 50 mM sodium acetate pH 6 buffer plus 200 mM NaCl. The purified protein is stored at 4°C or frozen at -80°C.

20 In addition to the above expression vector, the present invention further includes an expression vector comprising phage operator and promoter elements operatively linked to a polynucleotide of the present invention, called pHE4a. (ATCC Accession Number XXXXXX.) This vector contains: 1) a neomycinphosphotransferase gene as a selection marker, 2) an E. coli origin of replication, 3) a T5 phage promoter sequence,  
25 4) two lac operator sequences, 5) a Shine-Delgarno sequence, and 6) the lactose operon repressor gene (*lacIq*). The origin of replication (*oriC*) is derived from pUC19 (LTI, Gaithersburg, MD). The promoter sequence and operator sequences are made synthetically.

DNA can be inserted into the pHEa by restricting the vector with NdeI and  
30 XbaI, BamHI, XhoI, or Asp718, running the restricted product on a gel, and isolating the larger fragment (the stuffer fragment should be about 310 base pairs). The DNA insert is generated according to the PCR protocol described in Example 1, using PCR primers having restriction sites for NdeI (5' primer) and XbaI, BamHI, XhoI, or Asp718 (3' primer). The PCR insert is gel purified and restricted with compatible  
35 enzymes. The insert and vector are ligated according to standard protocols.

The engineered vector could easily be substituted in the above protocol to express protein in a bacterial system.

**Example 6: Purification of a Polypeptide from an Inclusion Body**

- 5       The following alternative method can be used to purify a polypeptide expressed in *E. coli* when it is present in the form of inclusion bodies. Unless otherwise specified, all of the following steps are conducted at 4-10°C.

      Upon completion of the production phase of the *E. coli* fermentation, the cell culture is cooled to 4-10°C and the cells harvested by continuous centrifugation at  
10   15,000 rpm (Heraeus Sepatech). On the basis of the expected yield of protein per unit weight of cell paste and the amount of purified protein required, an appropriate amount of cell paste, by weight, is suspended in a buffer solution containing 100 mM Tris, 50 mM EDTA, pH 7.4. The cells are dispersed to a homogeneous suspension using a high shear mixer.

15       The cells are then lysed by passing the solution through a microfluidizer (Microfluidics, Corp. or APV Gaulin, Inc.) twice at 4000-6000 psi. The homogenate is then mixed with NaCl solution to a final concentration of 0.5 M NaCl, followed by centrifugation at 7000 xg for 15 min. The resultant pellet is washed again using 0.5M NaCl, 100 mM Tris, 50 mM EDTA, pH 7.4.

20       The resulting washed inclusion bodies are solubilized with 1.5 M guanidine hydrochloride (GuHCl) for 2-4 hours. After 7000 xg centrifugation for 15 min., the pellet is discarded and the polypeptide containing supernatant is incubated at 4°C overnight to allow further GuHCl extraction.

      Following high speed centrifugation (30,000 xg) to remove insoluble particles,  
25   the GuHCl solubilized protein is refolded by quickly mixing the GuHCl extract with 20 volumes of buffer containing 50 mM sodium, pH 4.5, 150 mM NaCl, 2 mM EDTA by vigorous stirring. The refolded diluted protein solution is kept at 4°C without mixing for 12 hours prior to further purification steps.

      To clarify the refolded polypeptide solution, a previously prepared tangential  
30   filtration unit equipped with 0.16 µm membrane filter with appropriate surface area (e.g., Filtron), equilibrated with 40 mM sodium acetate, pH 6.0 is employed. The filtered sample is loaded onto a cation exchange resin (e.g., Poros HS-50, Perseptive Biosystems). The column is washed with 40 mM sodium acetate, pH 6.0 and eluted with 250 mM, 500 mM, 1000 mM, and 1500 mM NaCl in the same buffer, in a

stepwise manner. The absorbance at 280 nm of the effluent is continuously monitored. Fractions are collected and further analyzed by SDS-PAGE.

Fractions containing the polypeptide are then pooled and mixed with 4 volumes of water. The diluted sample is then loaded onto a previously prepared set of tandem  
5 columns of strong anion (Poros HQ-50, Perseptive Biosystems) and weak anion (Poros CM-20, Perseptive Biosystems) exchange resins. The columns are equilibrated with 40 mM sodium acetate, pH 6.0. Both columns are washed with 40 mM sodium acetate, pH 6.0, 200 mM NaCl. The CM-20 column is then eluted using a 10 column  
10 volume linear gradient ranging from 0.2 M NaCl, 50 mM sodium acetate, pH 6.0 to 1.0 M NaCl, 50 mM sodium acetate, pH 6.5. Fractions are collected under constant  $A_{280}$  monitoring of the effluent. Fractions containing the polypeptide (determined, for instance, by 16% SDS-PAGE) are then pooled.

The resultant polypeptide should exhibit greater than 95% purity after the above refolding and purification steps. No major contaminant bands should be observed from  
15 Commassie blue stained 16% SDS-PAGE gel when 5  $\mu$ g of purified protein is loaded. The purified protein can also be tested for endotoxin/LPS contamination, and typically the LPS content is less than 0.1 ng/ml according to LAL assays.

#### **Example 7: Cloning and Expression of a Polypeptide in a Baculovirus Expression System**

20

In this example, the plasmid shuttle vector pA2 is used to insert a polynucleotide into a baculovirus to express a polypeptide. This expression vector contains the strong polyhedrin promoter of the *Autographa californica* nuclear polyhedrosis virus (AcMNPV) followed by convenient restriction sites such as BamHI, Xba I and  
25 Asp718. The polyadenylation site of the simian virus 40 ("SV40") is used for efficient polyadenylation. For easy selection of recombinant virus, the plasmid contains the beta-galactosidase gene from *E. coli* under control of a weak *Drosophila* promoter in the same orientation, followed by the polyadenylation signal of the polyhedrin gene. The inserted genes are flanked on both sides by viral sequences for cell-mediated  
30 homologous recombination with wild-type viral DNA to generate a viable virus that express the cloned polynucleotide.

Many other baculovirus vectors can be used in place of the vector above, such as pAc373, pVL941, and pAcIM1, as one skilled in the art would readily appreciate, as long as the construct provides appropriately located signals for transcription,  
35 translation, secretion and the like, including a signal peptide and an in-frame AUG as

required. Such vectors are described, for instance, in Luckow et al., *Virology* 170:31-39 (1989).

Specifically, the cDNA sequence contained in the deposited clone, including the AUG initiation codon and the naturally associated leader sequence identified in Table 1, is amplified using the PCR protocol described in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the pA2 vector does not need a second signal peptide. Alternatively, the vector can be modified (pA2 GP) to include a baculovirus leader sequence, using the standard methods described in Summers et al., "A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures," Texas Agricultural Experimental Station Bulletin No. 1555 (1987).

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("GeneClean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

The plasmid is digested with the corresponding restriction enzymes and optionally, can be dephosphorylated using calf intestinal phosphatase, using routine procedures known in the art. The DNA is then isolated from a 1% agarose gel using a commercially available kit ("GeneClean" BIO 101 Inc., La Jolla, Ca.).

The fragment and the dephosphorylated plasmid are ligated together with T4 DNA ligase. *E. coli* HB101 or other suitable *E. coli* hosts such as XL-1 Blue (Stratagene Cloning Systems, La Jolla, CA) cells are transformed with the ligation mixture and spread on culture plates. Bacteria containing the plasmid are identified by digesting DNA from individual colonies and analyzing the digestion product by gel electrophoresis. The sequence of the cloned fragment is confirmed by DNA sequencing.

Five  $\mu$ g of a plasmid containing the polynucleotide is co-transfected with 1.0  $\mu$ g of a commercially available linearized baculovirus DNA ("BaculoGold™ baculovirus DNA", Pharmingen, San Diego, CA), using the lipofection method described by Felgner et al., *Proc. Natl. Acad. Sci. USA* 84:7413-7417 (1987). One  $\mu$ g of BaculoGold™ virus DNA and 5  $\mu$ g of the plasmid are mixed in a sterile well of a microtiter plate containing 50  $\mu$ l of serum-free Grace's medium (Life Technologies Inc., Gaithersburg, MD). Afterwards, 10  $\mu$ l Lipofectin plus 90  $\mu$ l Grace's medium are added, mixed and incubated for 15 minutes at room temperature. Then the transfection mixture is added drop-wise to Sf9 insect cells (ATCC CRL 1711) seeded in a 35 mm tissue culture plate with 1 ml Grace's medium without serum. The plate is then incubated for 5 hours at 27° C. The transfection solution is then removed from the plate and 1 ml of Grace's insect medium supplemented with 10% fetal calf serum is added. Cultivation is then continued at 27° C for four days.

After four days the supernatant is collected and a plaque assay is performed, as described by Summers and Smith, *supra*. An agarose gel with "Blue Gal" (Life Technologies Inc., Gaithersburg) is used to allow easy identification and isolation of gal-expressing clones, which produce blue-stained plaques. (A detailed description of a  
5 "plaque assay" of this type can also be found in the user's guide for insect cell culture and baculovirology distributed by Life Technologies Inc., Gaithersburg, page 9-10.) After appropriate incubation, blue stained plaques are picked with the tip of a micropipettor (e.g., Eppendorf). The agar containing the recombinant viruses is then resuspended in a microcentrifuge tube containing 200  $\mu$ l of Grace's medium and the  
10 suspension containing the recombinant baculovirus is used to infect Sf9 cells seeded in 35 mm dishes. Four days later the supernatants of these culture dishes are harvested and then they are stored at 4° C.

To verify the expression of the polypeptide, Sf9 cells are grown in Grace's medium supplemented with 10% heat-inactivated FBS. The cells are infected with the  
15 recombinant baculovirus containing the polynucleotide at a multiplicity of infection ("MOI") of about 2. If radiolabeled proteins are desired, 6 hours later the medium is removed and is replaced with SF900 II medium minus methionine and cysteine (available from Life Technologies Inc., Rockville, MD). After 42 hours, 5  $\mu$ Ci of  $^{35}$ S-methionine and 5  $\mu$ Ci  $^{35}$ S-cysteine (available from Amersham) are added. The cells are  
20 further incubated for 16 hours and then are harvested by centrifugation. The proteins in the supernatant as well as the intracellular proteins are analyzed by SDS-PAGE followed by autoradiography (if radiolabeled).

Microsequencing of the amino acid sequence of the amino terminus of purified protein may be used to determine the amino terminal sequence of the produced protein.

#### 25 **Example 8: Expression of a Polypeptide in Mammalian Cells**

The polypeptide of the present invention can be expressed in a mammalian cell. A typical mammalian expression vector contains a promoter element, which mediates the initiation of transcription of mRNA, a protein coding sequence, and signals required for the termination of transcription and polyadenylation of the transcript. Additional  
30 elements include enhancers, Kozak sequences and intervening sequences flanked by donor and acceptor sites for RNA splicing. Highly efficient transcription is achieved with the early and late promoters from SV40, the long terminal repeats (LTRs) from Retroviruses, e.g., RSV, HTLVI, HIVI and the early promoter of the cytomegalovirus (CMV). However, cellular elements can also be used (e.g., the human actin promoter).

35 Suitable expression vectors for use in practicing the present invention include, for example, vectors such as pSVL and pMSG (Pharmacia, Uppsala, Sweden),

pRSVcat (ATCC 37152), pSV2dhfr (ATCC 37146), pBC12MI (ATCC 67109), pCMVSPORT 2.0, and pCMVSPORT 3.0. Mammalian host cells that could be used include, human HeLa, 293, H9 and Jurkat cells, mouse NIH3T3 and C127 cells, Cos 1, Cos 7 and CV1, quail QC1-3 cells, mouse L cells and Chinese hamster ovary (CHO) cells.

Alternatively, the polypeptide can be expressed in stable cell lines containing the polynucleotide integrated into a chromosome. The co-transfection with a selectable marker such as dhfr, gpt, neomycin, hygromycin allows the identification and isolation of the transfected cells.

The transfected gene can also be amplified to express large amounts of the encoded protein. The DHFR (dihydrofolate reductase) marker is useful in developing cell lines that carry several hundred or even several thousand copies of the gene of interest. (See, e.g., Alt, F. W., et al., *J. Biol. Chem.* 253:1357-1370 (1978); Hamlin, J. L. and Ma, C., *Biochem. et Biophys. Acta*, 1097:107-143 (1990); Page, M. J. and Sydenham, M. A., *Biotechnology* 9:64-68 (1991).) Another useful selection marker is the enzyme glutamine synthase (GS) (Murphy et al., *Biochem J.* 227:277-279 (1991); Bebbington et al., *Bio/Technology* 10:169-175 (1992). Using these markers, the mammalian cells are grown in selective medium and the cells with the highest resistance are selected. These cell lines contain the amplified gene(s) integrated into a chromosome. Chinese hamster ovary (CHO) and NSO cells are often used for the production of proteins.

Derivatives of the plasmid pSV2-dhfr (ATCC Accession No. 37146), the expression vectors pC4 (ATCC Accession No. 209646) and pC6 (ATCC Accession No. 209647) contain the strong promoter (LTR) of the Rous Sarcoma Virus (Cullen et al., *Molecular and Cellular Biology*, 438-447 (March, 1985)) plus a fragment of the CMV-enhancer (Boshart et al., *Cell* 41:521-530 (1985).) Multiple cloning sites, e.g., with the restriction enzyme cleavage sites BamHI, XbaI and Asp718, facilitate the cloning of the gene of interest. The vectors also contain the 3' intron, the polyadenylation and termination signal of the rat preproinsulin gene, and the mouse DHFR gene under control of the SV40 early promoter.

Specifically, the plasmid pC6, for example, is digested with appropriate restriction enzymes and then dephosphorylated using calf intestinal phosphates by procedures known in the art. The vector is then isolated from a 1% agarose gel.

A polynucleotide of the present invention is amplified according to the protocol outlined in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the vector does not need a second signal peptide. Alternatively, if the



naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("GeneClean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested  
5 with appropriate restriction enzymes and again purified on a 1% agarose gel.

The amplified fragment is then digested with the same restriction enzyme and purified on a 1% agarose gel. The isolated fragment and the dephosphorylated vector are then ligated with T4 DNA ligase. *E. coli* HB101 or XL-1 Blue cells are then transformed and bacteria are identified that contain the fragment inserted into plasmid  
10 pC6 using, for instance, restriction enzyme analysis.

Chinese hamster ovary cells lacking an active DHFR gene is used for transfection. Five  $\mu$ g of the expression plasmid pC6 is cotransfected with 0.5  $\mu$ g of the plasmid pSVneo using lipofectin (Felgner et al., *supra*). The plasmid pSV2-neo contains a dominant selectable marker, the *neo* gene from Tn5 encoding an enzyme that  
15 confers resistance to a group of antibiotics including G418. The cells are seeded in alpha minus MEM supplemented with 1 mg/ml G418. After 2 days, the cells are trypsinized and seeded in hybridoma cloning plates (Greiner, Germany) in alpha minus MEM supplemented with 10, 25, or 50 ng/ml of methotrexate plus 1 mg/ml G418. After about 10-14 days single clones are trypsinized and then seeded in 6-well petri  
20 dishes or 10 ml flasks using different concentrations of methotrexate (50 nM, 100 nM, 200 nM, 400 nM, 800 nM). Clones growing at the highest concentrations of methotrexate are then transferred to new 6-well plates containing even higher concentrations of methotrexate (1  $\mu$ M, 2  $\mu$ M, 5  $\mu$ M, 10 mM, 20 mM). The same procedure is repeated until clones are obtained which grow at a concentration of 100 -  
25 200  $\mu$ M. Expression of the desired gene product is analyzed, for instance, by SDS-PAGE and Western blot or by reversed phase HPLC analysis.

#### **Example 9: Protein Fusions**

The polypeptides of the present invention are preferably fused to other proteins.  
30 These fusion proteins can be used for a variety of applications. For example, fusion of the present polypeptides to His-tag, HA-tag, protein A, IgG domains, and maltose binding protein facilitates purification. (See Example 5; see also EP A 394,827; Traunecker, et al., Nature 331:84-86 (1988).) Similarly, fusion to IgG-1, IgG-3, and albumin increases the halflife time in vivo. Nuclear localization signals fused to the  
35 polypeptides of the present invention can target the protein to a specific subcellular localization, while covalent heterodimer or homodimers can increase or decrease the activity of a fusion protein. Fusion proteins can also create chimeric molecules having

more than one function. Finally, fusion proteins can increase solubility and/or stability of the fused protein compared to the non-fused protein. All of the types of fusion proteins described above can be made by modifying the following protocol, which outlines the fusion of a polypeptide to an IgG molecule, or the protocol described in

5 Example 5.

Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector.

10 For example, if pC4 (Accession No.209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 1, is ligated into this BamHI site. Note that  
15 the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced.

If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a  
20 heterologous signal sequence. (See, e.g., WO 96/34891.)

Human IgG Fc region:

GGGATCCGGAGCCCAAATCTTCTGACAAACTCACACATGCCCACCGTGCC  
CAGCACCTGAATTCGAGGGTGCACCGTCAGTCTTCCTCTTCCCCCAAACC  
25 CAAGGACACCCTCATGATCTCCCGGACTCCTGAGGTCACATGCGTGGTGGT  
GGACGTAAGCCACGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGGACG  
GCGTGGAGGTGCATAATGCCAAGACAAAGCCGCGGGAGGAGCAGTACAAC  
AGCACGTACCGTGTGGTCAGCGTCCTCACCGTCCTGCACCAGGACTGGCTG  
AATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCCCTCCCAACCCCC  
30 ATCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCCGAGAACCACAGGT  
GTACACCCTGCCCCCATCCCGGGATGAGCTGACCAAGAACCAGGTCAGCCT  
GACCTGCCTGGTCAAAGGCTTCTATCCAAGCGACATCGCCGTGGAGTGGGA  
GAGCAATGGGCAGCCGGAGAACAACACTACAAGACCACGCCTCCCGTGCTGG  
ACTCCGACGGCTCCTTCTTCCTCTACAGCAAGCTCACCGTGGACAAGAGCA  
35 GGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGTGATGCATGAGGCTCTGC  
ACAACCACTACACGCAGAAGAGCCTCTCCCTGTCTCCGGGTAAATGAGTGC  
GACGGCCGCGACTCTAGAGGAT (SEQ ID NO:1)

**Example 10: Production of an Antibody from a Polypeptide**

The antibodies of the present invention can be prepared by a variety of methods. (See, Current Protocols, Chapter 2.) For example, cells expressing a polypeptide of the present invention is administered to an animal to induce the production of sera containing polyclonal antibodies. In a preferred method, a preparation of the secreted protein is prepared and purified to render it substantially free of natural contaminants. Such a preparation is then introduced into an animal in order to produce polyclonal antisera of greater specific activity.

In the most preferred method, the antibodies of the present invention are monoclonal antibodies (or protein binding fragments thereof). Such monoclonal antibodies can be prepared using hybridoma technology. (Köhler et al., Nature 256:495 (1975); Köhler et al., Eur. J. Immunol. 6:511 (1976); Köhler et al., Eur. J. Immunol. 6:292 (1976); Hammerling et al., in: Monoclonal Antibodies and T-Cell Hybridomas, Elsevier, N.Y., pp. 563-681 (1981).) In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100 µg/ml of streptomycin.

The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP2O), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands et al. (Gastroenterology 80:225-232 (1981).) The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide.

Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody

whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein-specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies.

5 It will be appreciated that Fab and F(ab')<sub>2</sub> and other fragments of the antibodies of the present invention may be used according to the methods disclosed herein. Such fragments are typically produced by proteolytic cleavage, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab')<sub>2</sub> fragments). Alternatively, secreted protein-binding fragments can be produced through the application of  
10 recombinant DNA technology or through synthetic chemistry.

For in vivo use of antibodies in humans, it may be preferable to use "humanized" chimeric monoclonal antibodies. Such antibodies can be produced using genetic constructs derived from hybridoma cells producing the monoclonal antibodies described above. Methods for producing chimeric antibodies are known in the art.  
15 (See, for review, Morrison, Science 229:1202 (1985); Oi et al., BioTechniques 4:214 (1986); Cabilly et al., U.S. Patent No. 4,816,567; Taniguchi et al., EP 171496; Morrison et al., EP 173494; Neuberger et al., WO 8601533; Robinson et al., WO 8702671; Boulianne et al., Nature 312:643 (1984); Neuberger et al., Nature 314:268 (1985).)

20

#### **Example 11: Production Of Secreted Protein For High-Throughput Screening Assays**

The following protocol produces a supernatant containing a polypeptide to be tested. This supernatant can then be used in the Screening Assays described in  
25 Examples 13-20.

First, dilute Poly-D-Lysine (644 587 Boehringer-Mannheim) stock solution (1mg/ml in PBS) 1:20 in PBS (w/o calcium or magnesium 17-516F Biowhittaker) for a working solution of 50ug/ml. Add 200 ul of this solution to each well (24 well plates) and incubate at RT for 20 minutes. Be sure to distribute the solution over each well  
30 (note: a 12-channel pipetter may be used with tips on every other channel). Aspirate off the Poly-D-Lysine solution and rinse with 1ml PBS (Phosphate Buffered Saline). The PBS should remain in the well until just prior to plating the cells and plates may be poly-lysine coated in advance for up to two weeks.

Plate 293T cells (do not carry cells past P+20) at  $2 \times 10^5$  cells/well in .5ml  
35 DMEM(Dulbecco's Modified Eagle Medium)(with 4.5 G/L glucose and L-glutamine (12-604F Biowhittaker))/10% heat inactivated FBS(14-503F Biowhittaker)/1x Penstrep(17-602E Biowhittaker). Let the cells grow overnight.

The next day, mix together in a sterile solution basin: 300 ul Lipofectamine (18324-012 Gibco/BRL) and 5ml Optimem I (31985070 Gibco/BRL)/96-well plate. With a small volume multi-channel pipetter, aliquot approximately 2ug of an expression vector containing a polynucleotide insert, produced by the methods described in

- 5 Examples 8 or 9, into an appropriately labeled 96-well round bottom plate. With a multi-channel pipetter, add 50ul of the Lipofectamine/Optimem I mixture to each well. Pipette up and down gently to mix. Incubate at RT 15-45 minutes. After about 20 minutes, use a multi-channel pipetter to add 150ul Optimem I to each well. As a control, one plate of vector DNA lacking an insert should be transfected with each set of
- 10 transfections.

- Preferably, the transfection should be performed by tag-teaming the following tasks. By tag-teaming, hands on time is cut in half, and the cells do not spend too much time on PBS. First, person A aspirates off the media from four 24-well plates of cells, and then person B rinses each well with .5-1ml PBS. Person A then aspirates off
- 15 PBS rinse, and person B, using a 12-channel pipetter with tips on every other channel, adds the 200ul of DNA/Lipofectamine/Optimem I complex to the odd wells first, then to the even wells, to each row on the 24-well plates. Incubate at 37°C for 6 hours.

- While cells are incubating, prepare appropriate media, either 1%BSA in DMEM with 1x penstrep, or CHO-5 media (see below) with 2mm glutamine and 1x penstrep.
- 20 (BSA (81-068-3 Bayer) 100gm dissolved in 1L DMEM for a 10% BSA stock solution). Filter the media and collect 50 ul for endotoxin assay in 15ml polystyrene conical.

- The transfection reaction is terminated, preferably by tag-teaming, at the end of the incubation period. Person A aspirates off the transfection media, while person B
- 25 adds 1.5ml appropriate media to each well. Incubate at 37°C for 45 or 72 hours depending on the media used: 1%BSA for 45 hours or CHO-5 for 72 hours.

On day four, using a 300ul multichannel pipetter, aliquot 600ul in one 1ml deep well plate and the remaining supernatant into a 2ml deep well. The supernatants from each well can then be used in the assays described in Examples 13-20.

- 30 It is specifically understood that when activity is obtained in any of the assays described below using a supernatant, the activity originates from either the polypeptide directly (e.g., as a secreted protein) or by the polypeptide inducing expression of other proteins, which are then secreted into the supernatant. Thus, the invention further provides a method of identifying the protein in the supernatant characterized by an
- 35 activity in a particular assay.

**HGS-CHO-5 medium formulation:****Inorganic Salts**

CaCl <sub>2</sub> (anhyd)	116.6 mg/L
CuSO <sub>4</sub> ·5H <sub>2</sub> O	0.00130
Fe(NO <sub>3</sub> ) <sub>3</sub> ·9H <sub>2</sub> O	0.050
FeSO <sub>4</sub> ·7H <sub>2</sub> O	0.417
KCl	311.80
MgCl <sub>2</sub>	28.64
MgSO <sub>4</sub>	48.84
NaCl	6995.50
NaHCO <sub>3</sub>	2400.0
NaH <sub>2</sub> PO <sub>4</sub> ·H <sub>2</sub> O	62.50
Na <sub>2</sub> HPO <sub>4</sub>	71.02
ZnSO <sub>4</sub> ·7H <sub>2</sub> O	.4320

**5 Lipids**

Arachidonic Acid	.002 mg/L
Cholesterol	1.022
DL-alpha-Tocopherol-Acetate	.070
Linoleic Acid	0.0520
Linolenic Acid	0.010
Myristic Acid	0.010
Oleic Acid	0.010
Palmitric Acid	0.010
Palmitic Acid	0.010
Pluronic F-68	100
Stearic Acid	0.010
Tween 80	2.20

**Carbon Source**

D-Glucose	4551 mg/L
-----------	-----------

**Amino Acids**

L- Alanine	130.85 mg/ml
L-Arginine-HCL	147.50
L-Asparagine-H <sub>2</sub> O	7.50
L-Aspartic Acid	6.65
L-Cystine-2HCL-H <sub>2</sub> O	29.56
L-Cystine-2HCL	31.29
L-Glutamic Acid	7.35
L-Glutamine	365.0
Glycine	18.75
L-Histidine-HCL-	52.48

H <sub>2</sub> O	
L-Isoleucine	106.97
L-Leucine	111.45
L-Lysine HCL	163.75
L-Methionine	32.34
L-Phenylalanine	68.48
L-Proline	40.0
L-Serine	26.25
L-Threonine	101.05
L-Tryptophan	19.22
L-Tyrosine-2Na-2H <sub>2</sub> O	91.79
L-Valine	99.65

### Vitamins

Biotin	0.0035 mg/L
D-Ca Pantothenate	3.24
Choline Chloride	11.78
Folic Acid	4.65
i-Inositol	15.60
Niacinamide	3.02
Pyridoxal HCL	3.00
Pyridoxine HCL	0.031
Riboflavin	0.319
Thiamine HCL	3.17
Thymidine	0.365
Vitamin B <sub>12</sub>	0.680

### Other Components

HEPES Buffer	25 mM
Na Hypoxanthine	2.39 mg/L
Lipoic Acid	0.105
Sodium Putrescine-2HCL	0.081
Sodium Pyruvate	55.0
Sodium Selenite	0.0067
Ethanolamine	20uM
Ferric Citrate	0.122
Methyl-B-Cyclodextrin complexed with Linoleic Acid	41.70
Methyl-B-Cyclodextrin complexed with Oleic Acid	33.33
Methyl-B-Cyclodextrin complexed with Retinal Acetate	10

5

*Adjust osmolarity to 327 mOsm*

**Example 12: Construction of GAS Reporter Construct**

One signal transduction pathway involved in the differentiation and proliferation of cells is called the Jaks-STATs pathway. Activated proteins in the Jaks-STATs pathway bind to gamma activation site "GAS" elements or interferon-sensitive responsive element ("ISRE"), located in the promoter of many genes. The binding of a protein to these elements alter the expression of the associated gene.

GAS and ISRE elements are recognized by a class of transcription factors called Signal Transducers and Activators of Transcription, or "STATs." There are six members of the STATs family. Stat1 and Stat3 are present in many cell types, as is Stat2 (as response to IFN-alpha is widespread). Stat4 is more restricted and is not in many cell types though it has been found in T helper class I, cells after treatment with IL-12. Stat5 was originally called mammary growth factor, but has been found at higher concentrations in other cells including myeloid cells. It can be activated in tissue culture cells by many cytokines.

The STATs are activated to translocate from the cytoplasm to the nucleus upon tyrosine phosphorylation by a set of kinases known as the Janus Kinase ("Jaks") family. Jaks represent a distinct family of soluble tyrosine kinases and include Tyk2, Jak1, Jak2, and Jak3. These kinases display significant sequence similarity and are generally catalytically inactive in resting cells.

The Jaks are activated by a wide range of receptors summarized in the Table below. (Adapted from review by Schidler and Darnell, Ann. Rev. Biochem. 64:621-51 (1995).) A cytokine receptor family, capable of activating Jaks, is divided into two groups: (a) Class 1 includes receptors for IL-2, IL-3, IL-4, IL-6, IL-7, IL-9, IL-11, IL-12, IL-15, Epo, PRL, GH, G-CSF, GM-CSF, LIF, CNTF, and thrombopoietin; and (b) Class 2 includes IFN-a, IFN-g, and IL-10. The Class 1 receptors share a conserved cysteine motif (a set of four conserved cysteines and one tryptophan) and a WSXWS motif (a membrane proximal region encoding Trp-Ser-Xxx-Trp-Ser (SEQ ID NO:2)).

Thus, on binding of a ligand to a receptor, Jaks are activated, which in turn activate STATs, which then translocate and bind to GAS elements. This entire process is encompassed in the Jaks-STATs signal transduction pathway.

Therefore, activation of the Jaks-STATs pathway, reflected by the binding of the GAS or the ISRE element, can be used to indicate proteins involved in the proliferation and differentiation of cells. For example, growth factors and cytokines are known to activate the Jaks-STATs pathway. (See Table below.) Thus, by using GAS elements linked to reporter molecules, activators of the Jaks-STATs pathway can be identified.



	<u>ISRE</u> <u>Ligand</u>	<u>JAKs</u>				<u>STATs</u>	<u>GAS(elements) or</u>
		<u>tyk2</u>	<u>Jak1</u>	<u>Jak2</u>	<u>Jak3</u>		
5	<u>IFN family</u>						
	IFN-a/B	+	+	-	-	1,2,3	ISRE
	IFN-g		+	+	-	1	GAS
	(IRF1>Lys6>IFP)						
	IL-10	+	?	?	-	1,3	
10	<u>gp130 family</u>						
	IL-6 (Pleiotrohic)	+	+	+	?	1,3	GAS
	(IRF1>Lys6>IFP)						
	IL-11(Pleiotrohic)	?	+	?	?	1,3	
15	OnM(Pleiotrohic)	?	+	+	?	1,3	
	LIF(Pleiotrohic)	?	+	+	?	1,3	
	CNTF(Pleiotrohic)	-/+	+	+	?	1,3	
	G-CSF(Pleiotrohic)	?	+	?	?	1,3	
	IL-12(Pleiotrohic)	+	-	+	+	1,3	
20	<u>g-C family</u>						
	IL-2 (lymphocytes)	-	+	-	+	1,3,5	GAS
	IL-4 (lymph/myeloid)	-	+	-	+	6	GAS (IRF1 = IFP
	>>Ly6)(IgH)						
25	IL-7 (lymphocytes)	-	+	-	+	5	GAS
	IL-9 (lymphocytes)	-	+	-	+	5	GAS
	IL-13 (lymphocyte)	-	+	?	?	6	GAS
	IL-15	?	+	?	+	5	GAS
30	<u>gp140 family</u>						
	IL-3 (myeloid)	-	-	+	-	5	GAS
	(IRF1>IFP>>Ly6)						
	IL-5 (myeloid)	-	-	+	-	5	GAS
	GM-CSF (myeloid)	-	-	+	-	5	GAS
35	<u>Growth hormone family</u>						
	GH	?	-	+	-	5	
	PRL	?	+/-	+	-	1,3,5	
	EPO	?	-	+	-	5	
40	CAS>IRF1=IFP>>Ly6)						GAS(B-
	<u>Receptor Tyrosine Kinases</u>						
	EGF	?	+	+	-	1,3	GAS (IRF1)
45	PDGF	?	+	+	-	1,3	
	CSF-1	?	+	+	-	1,3	GAS (not IRF1)

To construct a synthetic GAS containing promoter element, which is used in the Biological Assays described in Examples 13-14, a PCR based strategy is employed to generate a GAS-SV40 promoter sequence. The 5' primer contains four tandem copies of the GAS binding site found in the IRF1 promoter and previously demonstrated to bind STATs upon induction with a range of cytokines (Rothman et al., Immunity 1:457-468 (1994).), although other GAS or ISRE elements can be used instead. The 5' primer also contains 18bp of sequence complementary to the SV40 early promoter sequence and is flanked with an XhoI site. The sequence of the 5' primer is:  
5':GCGCCTCGAGATTTCCTCCGAAATCTAGATTTCCTCCGAAATGATTTCCTCCG  
10 AAATGATTTCCTCCGAAATATCTGCCATCTCAATTAG:3' (SEQ ID NO:3)

The downstream primer is complementary to the SV40 promoter and is flanked with a Hind III site: 5':GCGGCAAGCTTTTGTCAAAGCCTAGGC:3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the B-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI/Hind III and subcloned into BLSK2-. (Stratagene.) Sequencing with forward and reverse primers confirms that the insert contains the following sequence:  
5':CTCGAGATTTCCTCCGAAATCTAGATTTCCTCCGAAATGATTTCCTCCGAAATG  
20 ATTTTCCTCCGAAATATCTGCCATCTCAATTAGTCAGCAACCATAGTCCCGCCC  
CTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGC  
CCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGC  
CTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTT  
TGCAAAAAGCTT:3' (SEQ ID NO:5)

25 With this GAS promoter element linked to the SV40 promoter, a GAS:SEAP2 reporter construct is next engineered. Here, the reporter molecule is a secreted alkaline phosphatase, or "SEAP." Clearly, however, any reporter molecule can be instead of SEAP, in this or in any of the other Examples. Well known reporter molecules that can be used instead of SEAP include chloramphenicol acetyltransferase (CAT), luciferase, alkaline phosphatase, B-galactosidase, green fluorescent protein (GFP), or any protein detectable by an antibody.

The above sequence confirmed synthetic GAS-SV40 promoter element is subcloned into the pSEAP-Promoter vector obtained from Clontech using HindIII and XhoI, effectively replacing the SV40 promoter with the amplified GAS:SV40 promoter element, to create the GAS-SEAP vector. However, this vector does not contain a

neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

Thus, in order to generate mammalian stable cell lines expressing the GAS-SEAP reporter, the GAS-SEAP cassette is removed from the GAS-SEAP vector using  
5 SalI and NotI, and inserted into a backbone vector containing the neomycin resistance gene, such as pGFP-1 (Clontech), using these restriction sites in the multiple cloning site, to create the GAS-SEAP/Neo vector. Once this vector is transfected into mammalian cells, this vector can then be used as a reporter molecule for GAS binding as described in Examples 13-14.

10 Other constructs can be made using the above description and replacing GAS with a different promoter sequence. For example, construction of reporter molecules containing NFK-B and EGR promoter sequences are described in Examples 15 and 16. However, many other promoters can be substituted using the protocols described in these Examples. For instance, SRE, IL-2, NFAT, or Osteocalcin promoters can be  
15 substituted, alone or in combination (e.g., GAS/NF-KB/EGR, GAS/NF-KB, IL-2/NFAT, or NF-KB/GAS). Similarly, other cell lines can be used to test reporter construct activity, such as HELA (epithelial), HUVEC (endothelial), Reh (B-cell), Saos-2 (osteoblast), HUVAC (aortic), or Cardiomyocyte.

20 **Example 13: High-Throughput Screening Assay for T-cell Activity.**

The following protocol is used to assess T-cell activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate T-cells. T-cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS  
25 signal transduction pathway. The T-cell used in this assay is Jurkat T-cells (ATCC Accession No. TIB-152), although Molt-3 cells (ATCC Accession No. CRL-1552) and Molt-4 cells (ATCC Accession No. CRL-1582) cells can also be used.

Jurkat T-cells are lymphoblastic CD4+ Th1 helper cells. In order to generate stable cell lines, approximately 2 million Jurkat cells are transfected with the GAS-  
30 SEAP/neo vector using DMRIE-C (Life Technologies)(transfection procedure described below). The transfected cells are seeded to a density of approximately 20,000 cells per well and transfectants resistant to 1 mg/ml gentamicin selected. Resistant colonies are expanded and then tested for their response to increasing concentrations of interferon gamma. The dose response of a selected clone is demonstrated.

35 Specifically, the following protocol will yield sufficient cells for 75 wells containing 200 ul of cells. Thus, it is either scaled up, or performed in multiple to generate sufficient cells for multiple 96 well plates. Jurkat cells are maintained in RPMI

+ 10% serum with 1% Pen-Strep. Combine 2.5 mls of OPTI-MEM (Life Technologies) with 10 ug of plasmid DNA in a T25 flask. Add 2.5 ml OPTI-MEM containing 50 ul of DMRIE-C and incubate at room temperature for 15-45 mins.

5 During the incubation period, count cell concentration, spin down the required number of cells ( $10^7$  per transfection), and resuspend in OPTI-MEM to a final concentration of  $10^7$  cells/ml. Then add 1ml of  $1 \times 10^7$  cells in OPTI-MEM to T25 flask and incubate at 37°C for 6 hrs. After the incubation, add 10 ml of RPMI + 15% serum.

10 The Jurkat:GAS-SEAP stable reporter lines are maintained in RPMI + 10% serum, 1 mg/ml Gentamicin, and 1% Pen-Strep. These cells are treated with supernatants containing a polypeptide as produced by the protocol described in Example 11.

On the day of treatment with the supernatant, the cells should be washed and resuspended in fresh RPMI + 10% serum to a density of 500,000 cells per ml. The exact number of cells required will depend on the number of supernatants being screened. For one 96 well plate, approximately 10 million cells (for 10 plates, 100 million cells) are required.

Transfer the cells to a triangular reservoir boat, in order to dispense the cells into a 96 well dish, using a 12 channel pipette. Using a 12 channel pipette, transfer 200 ul of cells into each well (therefore adding 100,000 cells per well).

20 After all the plates have been seeded, 50 ul of the supernatants are transferred directly from the 96 well plate containing the supernatants into each well using a 12 channel pipette. In addition, a dose of exogenous interferon gamma (0.1, 1.0, 10 ng) is added to wells H9, H10, and H11 to serve as additional positive controls for the assay.

25 The 96 well dishes containing Jurkat cells treated with supernatants are placed in an incubator for 48 hrs (note: this time is variable between 48-72 hrs). 35 ul samples from each well are then transferred to an opaque 96 well plate using a 12 channel pipette. The opaque plates should be covered (using sellophane covers) and stored at -20°C until SEAP assays are performed according to Example 17. The plates containing the remaining treated cells are placed at 4°C and serve as a source of material for repeating the assay on a specific well if desired.

30 As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate Jurkat T cells. Over 30 fold induction is typically observed in the positive control wells.

**Example 14: High-Throughput Screening Assay Identifying Myeloid Activity**

The following protocol is used to assess myeloid activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate myeloid cells.

- 5 Myeloid cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The myeloid cell used in this assay is U937, a pre-monocyte cell line, although TF-1, HL60, or KG1 can be used.

- 10 To transiently transfect U937 cells with the GAS/SEAP/Neo construct produced in Example 12, a DEAE-Dextran method (Kharbanda et. al., 1994, Cell Growth & Differentiation, 5:259-265) is used. First, harvest  $2 \times 10^7$  U937 cells and wash with PBS. The U937 cells are usually grown in RPMI 1640 medium containing 10% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 mg/ml streptomycin.

- 15 Next, suspend the cells in 1 ml of 20 mM Tris-HCl (pH 7.4) buffer containing 0.5 mg/ml DEAE-Dextran, 8 ug GAS-SEAP2 plasmid DNA, 140 mM NaCl, 5 mM KCl, 375 uM  $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$ , 1 mM  $\text{MgCl}_2$ , and 675 uM  $\text{CaCl}_2$ . Incubate at 37°C for 45 min.

- 20 Wash the cells with RPMI 1640 medium containing 10% FBS and then resuspend in 10 ml complete medium and incubate at 37°C for 36 hr.

The GAS-SEAP/U937 stable cells are obtained by growing the cells in 400 ug/ml G418. The G418-free medium is used for routine growth but every one to two months, the cells should be re-grown in 400 ug/ml G418 for couple of passages.

- 25 These cells are tested by harvesting  $1 \times 10^8$  cells (this is enough for ten 96-well plates assay) and wash with PBS. Suspend the cells in 200 ml above described growth medium, with a final density of  $5 \times 10^5$  cells/ml. Plate 200 ul cells per well in the 96-well plate (or  $1 \times 10^5$  cells/well).

- 30 Add 50 ul of the supernatant prepared by the protocol described in Example 11. Incubate at 37°C for 48 to 72 hr. As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate U937 cells. Over 30 fold induction is typically observed in the positive control wells. SEAP assay the supernatant according to the protocol described in Example 17.

**Example 15: High-Throughput Screening Assay Identifying Neuronal Activity.**

When cells undergo differentiation and proliferation, a group of genes are activated through many different signal transduction pathways. One of these genes,  
5 EGR1 (early growth response gene 1), is induced in various tissues and cell types upon activation. The promoter of EGR1 is responsible for such induction. Using the EGR1 promoter linked to reporter molecules, activation of cells can be assessed.

Particularly, the following protocol is used to assess neuronal activity in PC12 cell lines. PC12 cells (rat pheochromocytoma cells) are known to proliferate and/or  
10 differentiate by activation with a number of mitogens, such as TPA (tetradecanoyl phorbol acetate), NGF (nerve growth factor), and EGF (epidermal growth factor). The EGR1 gene expression is activated during this treatment. Thus, by stably transfecting PC12 cells with a construct containing an EGR promoter linked to SEAP reporter, activation of PC12 cells can be assessed.

15 The EGR/SEAP reporter construct can be assembled by the following protocol. The EGR-1 promoter sequence (-633 to +1)(Sakamoto K et al., Oncogene 6:867-871 (1991)) can be PCR amplified from human genomic DNA using the following primers:

20 5' GCGCTCGAGGGATGACAGCGATAGAACCCCGG -3' (SEQ ID NO:6)  
5' GCGAAGCTTCGCGACTCCCCGGATCCGCCTC-3' (SEQ ID NO:7)

Using the GAS:SEAP/Neo vector produced in Example 12, EGR1 amplified product can then be inserted into this vector. Linearize the GAS:SEAP/Neo vector using restriction enzymes XhoI/HindIII, removing the GAS/SV40 stuffer. Restrict the  
25 EGR1 amplified product with these same enzymes. Ligate the vector and the EGR1 promoter.

To prepare 96 well-plates for cell culture, two mls of a coating solution (1:30 dilution of collagen type I (Upstate Biotech Inc. Cat#08-115) in 30% ethanol (filter sterilized)) is added per one 10 cm plate or 50 ml per well of the 96-well plate, and  
30 allowed to air dry for 2 hr.

PC12 cells are routinely grown in RPMI-1640 medium (Bio Whittaker) containing 10% horse serum (JRH BIOSCIENCES, Cat. # 12449-78P), 5% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 ug/ml streptomycin on a precoated 10 cm tissue culture dish. One to four split is done  
35 every three to four days. Cells are removed from the plates by scraping and resuspended with pipetting up and down for more than 15 times.

Transfect the EGR/SEAP/Neo construct into PC12 using the Lipofectamine protocol described in Example 11. EGR-SEAP/PC12 stable cells are obtained by growing the cells in 300 ug/ml G418. The G418-free medium is used for routine growth but every one to two months, the cells should be re-grown in 300 ug/ml G418 for couple of passages.

To assay for neuronal activity, a 10 cm plate with cells around 70 to 80% confluent is screened by removing the old medium. Wash the cells once with PBS (Phosphate buffered saline). Then starve the cells in low serum medium (RPMI-1640 containing 1% horse serum and 0.5% FBS with antibiotics) overnight.

The next morning, remove the medium and wash the cells with PBS. Scrape off the cells from the plate, suspend the cells well in 2 ml low serum medium. Count the cell number and add more low serum medium to reach final cell density as  $5 \times 10^5$  cells/ml.

Add 200 ul of the cell suspension to each well of 96-well plate (equivalent to  $1 \times 10^5$  cells/well). Add 50 ul supernatant produced by Example 11, 37°C for 48 to 72 hr. As a positive control, a growth factor known to activate PC12 cells through EGR can be used, such as 50 ng/ul of Neuronal Growth Factor (NGF). Over fifty-fold induction of SEAP is typically seen in the positive control wells. SEAP assay the supernatant according to Example 17.

#### **Example 16: High-Throughput Screening Assay for T-cell Activity**

NF- $\kappa$ B (Nuclear Factor  $\kappa$ B) is a transcription factor activated by a wide variety of agents including the inflammatory cytokines IL-1 and TNF, CD30 and CD40, lymphotoxin-alpha and lymphotoxin-beta, by exposure to LPS or thrombin, and by expression of certain viral gene products. As a transcription factor, NF- $\kappa$ B regulates the expression of genes involved in immune cell activation, control of apoptosis (NF- $\kappa$ B appears to shield cells from apoptosis), B and T-cell development, anti-viral and antimicrobial responses, and multiple stress responses.

In non-stimulated conditions, NF-  $\kappa$ B is retained in the cytoplasm with I- $\kappa$ B (Inhibitor  $\kappa$ B). However, upon stimulation, I-  $\kappa$ B is phosphorylated and degraded, causing NF-  $\kappa$ B to shuttle to the nucleus, thereby activating transcription of target genes. Target genes activated by NF-  $\kappa$ B include IL-2, IL-6, GM-CSF, ICAM-1 and class I MHC.

Due to its central role and ability to respond to a range of stimuli, reporter constructs utilizing the NF- $\kappa$ B promoter element are used to screen the supernatants produced in Example 11. Activators or inhibitors of NF- $\kappa$ B would be useful in treating diseases. For example, inhibitors of NF- $\kappa$ B could be used to treat those diseases  
5 related to the acute or chronic activation of NF- $\kappa$ B, such as rheumatoid arthritis.

To construct a vector containing the NF- $\kappa$ B promoter element, a PCR based strategy is employed. The upstream primer contains four tandem copies of the NF- $\kappa$ B binding site (GGGGACTTTCCC) (SEQ ID NO:8), 18 bp of sequence complementary to the 5' end of the SV40 early promoter sequence, and is flanked with an XhoI site:  
10 5':GCGGCCTCGAGGGGACTTTCCCGGGGACTTTCCGGGGACTTTCCGGGGAC  
TTTCCATCCTGCCATCTCAATTAG:3' (SEQ ID NO:9)

The downstream primer is complementary to the 3' end of the SV40 promoter and is flanked with a Hind III site:

5':GCGGCAAGCTTTTTGCAAAGCCTAGGC:3' (SEQ ID NO:4)  
15 PCR amplification is performed using the SV40 promoter template present in the pB-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI and Hind III and subcloned into BLSK2-. (Stratagene) Sequencing with the T7 and T3 primers confirms the insert contains the following sequence:

20 5':CTCGAGGGGACTTTCCCGGGGACTTTCCGGGGACTTTCCGGGGACTTTCC  
ATCTGCCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCCTAACTCCGCCCA  
TCCCGCCCCCTAACTCCGCCCAGTTCCGCCCATTTCTCCGCCCCATGGCTGACT  
AATTTTTTTTATTTATGCAGAGGCCGAGGCCGCTCGGCCTCTGAGCTATTC  
CAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTT:  
25 3' (SEQ ID NO:10)

Next, replace the SV40 minimal promoter element present in the pSEAP2- promoter plasmid (Clontech) with this NF- $\kappa$ B/SV40 fragment using XhoI and HindIII. However, this vector does not contain a neomycin resistance gene, and therefore, is not  
30 preferred for mammalian expression systems.

In order to generate stable mammalian cell lines, the NF- $\kappa$ B/SV40/SEAP cassette is removed from the above NF- $\kappa$ B/SEAP vector using restriction enzymes SalI and NotI, and inserted into a vector containing neomycin resistance. Particularly, the



NF- $\kappa$ B/SV40/SEAP cassette was inserted into pGFP-1 (Clontech), replacing the GFP gene, after restricting pGFP-1 with Sall and NotI.

- Once NF- $\kappa$ B/SV40/SEAP/Neo vector is created, stable Jurkat T-cells are created and maintained according to the protocol described in Example 13. Similarly, the method for assaying supernatants with these stable Jurkat T-cells is also described in Example 13. As a positive control, exogenous TNF alpha (0.1, 1, 10 ng) is added to wells H9, H10, and H11, with a 5-10 fold activation typically observed.

#### **Example 17: Assay for SEAP Activity**

- As a reporter molecule for the assays described in Examples 13-16, SEAP activity is assayed using the Tropix Phospho-light Kit (Cat. BP-400) according to the following general procedure. The Tropix Phospho-light Kit supplies the Dilution, Assay, and Reaction Buffers used below.

- Prime a dispenser with the 2.5x Dilution Buffer and dispense 15  $\mu$ l of 2.5x dilution buffer into Optiplates containing 35  $\mu$ l of a supernatant. Seal the plates with a plastic sealer and incubate at 65°C for 30 min. Separate the Optiplates to avoid uneven heating.

- Cool the samples to room temperature for 15 minutes. Empty the dispenser and prime with the Assay Buffer. Add 50  $\mu$ l Assay Buffer and incubate at room temperature 5 min. Empty the dispenser and prime with the Reaction Buffer (see the table below). Add 50  $\mu$ l Reaction Buffer and incubate at room temperature for 20 minutes. Since the intensity of the chemiluminescent signal is time dependent, and it takes about 10 minutes to read 5 plates on luminometer, one should treat 5 plates at each time and start the second set 10 minutes later.

- Read the relative light unit in the luminometer. Set H12 as blank, and print the results. An increase in chemiluminescence indicates reporter activity.

#### **Reaction Buffer Formulation:**

# of plates	Rxn buffer diluent (ml)	CSPD (ml)
10	60	3
11	65	3.25
12	70	3.5
13	75	3.75
14	80	4

15	85	4.25
16	90	4.5
17	95	4.75
18	100	5
19	105	5.25
20	110	5.5
21	115	5.75
22	120	6
23	125	6.25
24	130	6.5
25	135	6.75
26	140	7
27	145	7.25
28	150	7.5
29	155	7.75
30	160	8
31	165	8.25
32	170	8.5
33	175	8.75
34	180	9
35	185	9.25
36	190	9.5
37	195	9.75
38	200	10
39	205	10.25
40	210	10.5
41	215	10.75
42	220	11
43	225	11.25
44	230	11.5
45	235	11.75
46	240	12
47	245	12.25
48	250	12.5
49	255	12.75
50	260	13

---

**Example 18: High-Throughput Screening Assay Identifying Changes in Small Molecule Concentration and Membrane Permeability**

Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium, sodium, and pH, as well as alter membrane potential. These alterations can be measured in an assay to identify supernatants which bind to receptors of a particular cell. Although the following protocol describes an assay for calcium, this protocol can easily be modified to detect changes in potassium, sodium, pH, membrane potential, or any other small molecule which is detectable by a fluorescent probe.

The following assay uses Fluorometric Imaging Plate Reader ("FLIPR") to measure changes in fluorescent molecules (Molecular Probes) that bind small molecules. Clearly, any fluorescent molecule detecting a small molecule can be used instead of the calcium fluorescent molecule, fluo-3, used here.

For adherent cells, seed the cells at 10,000 -20,000 cells/well in a Co-star black 96-well plate with clear bottom. The plate is incubated in a CO<sub>2</sub> incubator for 20 hours. The adherent cells are washed two times in Biotek washer with 200 ul of HBSS (Hank's Balanced Salt Solution) leaving 100 ul of buffer after the final wash.

A stock solution of 1 mg/ml fluo-3 is made in 10% pluronic acid DMSO. To load the cells with fluo-3, 50 ul of 12 ug/ml fluo-3 is added to each well. The plate is incubated at 37°C in a CO<sub>2</sub> incubator for 60 min. The plate is washed four times in the Biotek washer with HBSS leaving 100 ul of buffer.

For non-adherent cells, the cells are spun down from culture media. Cells are re-suspended to 2-5x10<sup>6</sup> cells/ml with HBSS in a 50-ml conical tube. 4 ul of 1 mg/ml fluo-3 solution in 10% pluronic acid DMSO is added to each ml of cell suspension. The tube is then placed in a 37°C water bath for 30-60 min. The cells are washed twice with HBSS, resuspended to 1x10<sup>6</sup> cells/ml, and dispensed into a microplate, 100 ul/well. The plate is centrifuged at 1000 rpm for 5 min. The plate is then washed once in Denley CellWash with 200 ul, followed by an aspiration step to 100 ul final volume.

For a non-cell based assay, each well contains a fluorescent molecule, such as fluo-3. The supernatant is added to the well, and a change in fluorescence is detected.

To measure the fluorescence of intracellular calcium, the FLIPR is set for the following parameters: (1) System gain is 300-800 mW; (2) Exposure time is 0.4 second; (3) Camera F/stop is F/2; (4) Excitation is 488 nm; (5) Emission is 530 nm; and (6) Sample addition is 50 ul. Increased emission at 530 nm indicates an extracellular

signaling even which has resulted in an increase in the intracellular  $\text{Ca}^{++}$  concentration.

**Example 19: High-Throughput Screening Assay Identifying Tyrosine Kinase Activity**

The Protein Tyrosine Kinases (PTK) represent a diverse group of transmembrane and cytoplasmic kinases. Within the Receptor Protein Tyrosine Kinase (RPTK) group are receptors for a range of mitogenic and metabolic growth factors including the PDGF, FGF, EGF, NGF, HGF and Insulin receptor subfamilies. In addition there are a large family of RPTKs for which the corresponding ligand is unknown. Ligands for RPTKs include mainly secreted small proteins, but also membrane-bound and extracellular matrix proteins.

Activation of RPTK by ligands involves ligand-mediated receptor dimerization, resulting in transphosphorylation of the receptor subunits and activation of the cytoplasmic tyrosine kinases. The cytoplasmic tyrosine kinases include receptor associated tyrosine kinases of the src-family (e.g., src, yes, lck, lyn, fyn) and non-receptor linked and cytosolic protein tyrosine kinases, such as the Jak family, members of which mediate signal transduction triggered by the cytokine superfamily of receptors (e.g., the Interleukins, Interferons, GM-CSF, and Leptin).

Because of the wide range of known factors capable of stimulating tyrosine kinase activity, the identification of novel human secreted proteins capable of activating tyrosine kinase signal transduction pathways are of interest. Therefore, the following protocol is designed to identify those novel human secreted proteins capable of activating the tyrosine kinase signal transduction pathways.

Seed target cells (e.g., primary keratinocytes) at a density of approximately 25,000 cells per well in a 96 well Loprodyne Silent Screen Plates purchased from Nalge Nunc (Naperville, IL). The plates are sterilized with two 30 minute rinses with 100% ethanol, rinsed with water and dried overnight. Some plates are coated for 2 hr with 100 ml of cell culture grade type I collagen (50 mg/ml), gelatin (2%) or polylysine (50 mg/ml), all of which can be purchased from Sigma Chemicals (St. Louis, MO) or 10% Matrigel purchased from Becton Dickinson (Bedford, MA), or calf serum, rinsed with PBS and stored at 4°C. Cell growth on these plates is assayed by seeding 5,000 cells/well in growth medium and indirect quantitation of cell number through use of alamarBlue as described by the manufacturer Alamar Biosciences, Inc. (Sacramento, CA) after 48 hr. Falcon plate covers #3071 from Becton Dickinson (Bedford, MA) are

used to cover the Loprodyne Silent Screen Plates. Falcon Microtest III cell culture plates can also be used in some proliferation experiments.

To prepare extracts, A431 cells are seeded onto the nylon membranes of Loprodyne plates (20,000/200ml/well) and cultured overnight in complete medium.

- 5 Cells are quiesced by incubation in serum-free basal medium for 24 hr. After 5-20 minutes treatment with EGF (60ng/ml) or 50 ul of the supernatant produced in Example 11, the medium was removed and 100 ml of extraction buffer ((20 mM HEPES pH 7.5, 0.15 M NaCl, 1% Triton X-100, 0.1% SDS, 2 mM Na<sub>3</sub>VO<sub>4</sub>, 2 mM Na<sub>4</sub>P<sub>2</sub>O<sub>7</sub> and a cocktail of protease inhibitors (# 1836170) obtained from Boehringer Mannheim
- 10 (Indianapolis, IN) is added to each well and the plate is shaken on a rotating shaker for 5 minutes at 4°C. The plate is then placed in a vacuum transfer manifold and the extract filtered through the 0.45 mm membrane bottoms of each well using house vacuum. Extracts are collected in a 96-well catch/assay plate in the bottom of the vacuum manifold and immediately placed on ice. To obtain extracts clarified by centrifugation,
- 15 the content of each well, after detergent solubilization for 5 minutes, is removed and centrifuged for 15 minutes at 4°C at 16,000 x g.

Test the filtered extracts for levels of tyrosine kinase activity. Although many methods of detecting tyrosine kinase activity are known, one method is described here.

- Generally, the tyrosine kinase activity of a supernatant is evaluated by
- 20 determining its ability to phosphorylate a tyrosine residue on a specific substrate (a biotinylated peptide). Biotinylated peptides that can be used for this purpose include PSK1 (corresponding to amino acids 6-20 of the cell division kinase cdc2-p34) and PSK2 (corresponding to amino acids 1-17 of gastrin). Both peptides are substrates for a range of tyrosine kinases and are available from Boehringer Mannheim.

- 25 The tyrosine kinase reaction is set up by adding the following components in order. First, add 10ul of 5uM Biotinylated Peptide, then 10ul ATP/Mg<sub>2+</sub> (5mM ATP/50mM MgCl<sub>2</sub>), then 10ul of 5x Assay Buffer (40mM imidazole hydrochloride, pH7.3, 40 mM beta-glycerophosphate, 1mM EGTA, 100mM MgCl<sub>2</sub>, 5 mM MnCl<sub>2</sub>, 0.5 mg/ml BSA), then 5ul of Sodium Vanadate(1mM), and then 5ul of water. Mix the
- 30 components gently and preincubate the reaction mix at 30°C for 2 min. Initiate the reaction by adding 10ul of the control enzyme or the filtered supernatant.

The tyrosine kinase assay reaction is then terminated by adding 10 ul of 120mM EDTA and place the reactions on ice.

- Tyrosine kinase activity is determined by transferring 50 ul aliquot of reaction
- 35 mixture to a microtiter plate (MTP) module and incubating at 37°C for 20 min. This

allows the streptavidin coated 96 well plate to associate with the biotinylated peptide. Wash the MTP module with 300ul/well of PBS four times. Next add 75 ul of anti-phosphotyrosine antibody conjugated to horse radish peroxidase(anti-P-Tyr-POD(0.5u/ml)) to each well and incubate at 37°C for one hour. Wash the well as  
5 above.

Next add 100ul of peroxidase substrate solution (Boehringer Mannheim) and incubate at room temperature for at least 5 mins (up to 30 min). Measure the absorbance of the sample at 405 nm by using ELISA reader. The level of bound peroxidase activity is quantitated using an ELISA reader and reflects the level of  
10 tyrosine kinase activity.

**Example 20: High-Throughput Screening Assay Identifying Phosphorylation Activity**

As a potential alternative and/or complement to the assay of protein tyrosine  
15 kinase activity described in Example 19, an assay which detects activation (phosphorylation) of major intracellular signal transduction intermediates can also be used. For example, as described below one particular assay can detect tyrosine phosphorylation of the Erk-1 and Erk-2 kinases. However, phosphorylation of other molecules, such as Raf, JNK, p38 MAP, Map kinase kinase (MEK), MEK kinase,  
20 Src, Muscle specific kinase (MuSK), IRAK, Tec, and Janus, as well as any other phosphoserine, phosphotyrosine, or phosphothreonine molecule, can be detected by substituting these molecules for Erk-1 or Erk-2 in the following assay.

Specifically, assay plates are made by coating the wells of a 96-well ELISA plate with 0.1ml of protein G (1ug/ml) for 2 hr at room temp, (RT). The plates are then  
25 rinsed with PBS and blocked with 3% BSA/PBS for 1 hr at RT. The protein G plates are then treated with 2 commercial monoclonal antibodies (100ng/well) against Erk-1 and Erk-2 (1 hr at RT) (Santa Cruz Biotechnology). (To detect other molecules, this step can easily be modified by substituting a monoclonal antibody detecting any of the above described molecules.) After 3-5 rinses with PBS, the plates are stored at 4°C  
30 until use.

A431 cells are seeded at 20,000/well in a 96-well Loprodyne filterplate and cultured overnight in growth medium. The cells are then starved for 48 hr in basal medium (DMEM) and then treated with EGF (6ng/well) or 50 ul of the supernatants obtained in Example 11 for 5-20 minutes. The cells are then solubilized and extracts  
35 filtered directly into the assay plate.

After incubation with the extract for 1 hr at RT, the wells are again rinsed. As a positive control, a commercial preparation of MAP kinase (10ng/well) is used in place of A431 extract. Plates are then treated with a commercial polyclonal (rabbit) antibody (1ug/ml) which specifically recognizes the phosphorylated epitope of the Erk-1 and Erk-2 kinases (1 hr at RT). This antibody is biotinylated by standard procedures. The bound polyclonal antibody is then quantitated by successive incubations with Europium-streptavidin and Europium fluorescence enhancing reagent in the Wallac DELFIA instrument (time-resolved fluorescence). An increased fluorescent signal over background indicates a phosphorylation.

**Example 21: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide**

RNA isolated from entire families or individual patients presenting with a phenotype of interest (such as a disease) is be isolated. cDNA is then generated from these RNA samples using protocols known in the art. (See, Sambrook.) The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO:X. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky, D., et al., Science 252:706 (1991).

PCR products is then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies). The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations is then cloned and sequenced to validate the results of the direct sequencing.

PCR products is cloned into T-tailed vectors as described in Holton, T.A. and Graham, M.W., Nucleic Acids Research, 19:1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals is identified by mutations not present in unaffected individuals.

Genomic rearrangements are also observed as a method of determining alterations in a gene corresponding to a polynucleotide. Genomic clones isolated according to Example 2 are nick-translated with digoxigenindeoxy-uridine 5'-triphosphate (Boehringer Mannheim), and FISH performed as described in Johnson, Cg. et al., Methods Cell Biol. 35:73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C- and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera  
5 (Photometrics, Tucson, AZ) and variable excitation wavelength filters. (Johnson, Cv. et al., Genet. Anal. Tech. Appl., 8:75 (1991).) Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and  
10 translocations. These alterations are used as a diagnostic marker for an associated disease.

**Example 22: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample**

15 A polypeptide of the present invention can be detected in a biological sample, and if an increased or decreased level of the polypeptide is detected, this polypeptide is a marker for a particular phenotype. Methods of detection are numerous, and thus, it is understood that one skilled in the art can modify the following assay to fit their particular needs.

20 For example, antibody-sandwich ELISAs are used to detect soluble polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 ug/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described in Example 10. The wells are blocked so that non-specific binding of the  
25 polypeptide to the well is reduced.

The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbounded polypeptide.

30 Next, 50 ul of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbounded conjugate.

35 Add 75 ul of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution to each well and incubate 1 hour at room temperature. Measure the reaction by a microtiter plate reader. Prepare a standard curve, using serial dilutions of a control sample, and plot polypeptide concentration on



the X-axis (log scale) and fluorescence or absorbance of the Y-axis (linear scale).  
Interpolate the concentration of the polypeptide in the sample using the standard curve.

**Example 23: Formulating a Polypeptide**

5       The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for  
10       purposes herein is thus determined by such considerations.

      As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 µg/kg/day to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and  
15       most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 µg/kg/hour to about 50 µg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes  
20       and the interval following treatment for responses to occur appears to vary depending on the desired effect.

      Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracisternally, intravaginally, intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal  
25       patch), buccally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

30       The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semi-permeable polymer matrices in the form of shaped articles, e.g., films, or microcapsules. Sustained-release matrices include polylactides (U.S. Pat. No. 3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al.,  
35       Biopolymers 22:547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15:167-277 (1981), and R. Langer, Chem. Tech. 12:98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D-(-)-3-hydroxybutyric

acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE 3,218,121; Epstein et al., Proc. Natl. Acad. Sci. USA 82:3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77:4030-4034 (1980); EP 52,322; 5 EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U.S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

10 For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, i.e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation. For example, the 15 formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides.

Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the 20 carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that 25 enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e.g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or 30 immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrans; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, 35 poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of

about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile.

5 Sterility is readily accomplished by filtration through sterile filtration membranes (e.g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampoules or vials, as an aqueous solution or as a lyophilized  
10 formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1% (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or  
15 more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the  
20 present invention may be employed in conjunction with other therapeutic compounds.

#### **Example 24: Method of Treating Decreased Levels of the Polypeptide**

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by  
25 administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

30 For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 ug/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided in Example 23.

**Example 25: Method of Treating Increased Levels of the Polypeptide**

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

5 For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided in Example 23.

10

**Example 26: Method of Treatment Using Gene Therapy**

One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and  
15 separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e.g., Ham's F12 media, with 10% FBS,  
20 penicillin and streptomycin, is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks.

25 pMV-7 (Kirschmeier, P.T. et al., DNA, 7:219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified  
30 using PCR primers which correspond to the 5' and 3' end sequences respectively as set forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions  
35 appropriate for ligation of the two fragments. The ligation mixture is then used to

transform bacteria HB101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

5 The amphotropic pA317 or GP+am12 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

10 Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media. If the  
15 titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is being produced.

20 The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

It will be clear that the invention may be practiced otherwise than as particularly described in the foregoing description and examples. Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, are within the scope of the appended claims.

25 The entire disclosure of each document cited (including patents, patent applications, journal articles, abstracts, laboratory manuals, books, or other disclosures) in the Background of the Invention, Detailed Description, and Examples is hereby incorporated herein by reference.

## (1) GENERAL INFORMATION:

- 5 (i) APPLICANT: Human Genome Sciences, Inc. et al.
- (ii) TITLE OF INVENTION: 186 Human Secreted Proteins
- 10 (iii) NUMBER OF SEQUENCES: 644
- (iv) CORRESPONDENCE ADDRESS:
- (A) ADDRESSEE: Human Genome Sciences, Inc.
- 15 (B) STREET: 9410 Key West Avenue
- (C) CITY: Rockville
- (D) STATE: Maryland
- 20 (E) COUNTRY: USA
- (F) ZIP: 20850
- 25 (v) COMPUTER READABLE FORM:
- (A) MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
- 30 (B) COMPUTER: HP Vectra 486/33
- (C) OPERATING SYSTEM: MSDOS version 6.2
- 35 (D) SOFTWARE: ASCII Text
- (vi) CURRENT APPLICATION DATA:
- 40 (A) APPLICATION NUMBER:
- (B) FILING DATE: March 6, 1998
- 45 (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
- 50 (A) APPLICATION NUMBER:
- (B) FILING DATE:
- 55

## (viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: A. Anders Brookes, Esq.

(B) REGISTRATION NUMBER: 36,373

(C) REFERENCE/DOCKET NUMBER: PS002.PCT

## (vi) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (301) 309-8504

(B) TELEFAX: (301) 309-8439

## (2) INFORMATION FOR SEQ ID NO: 1:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 733 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GGGATCCGGA	GGCCAAATCT	TCTGACAAAA	CTCACACATG	CCCACCGTGC	CCAGCACCTG	60
AATTCGAGGG	TGCACCGTCA	GTCTTCCTCT	TCCCCCCTAA	ACCCAAGGAC	ACCCTCATGA	120
TCTCCCGGAC	TCCTGAGGTC	ACATGCGTGG	TGGTGGACGT	AAGCCACGAA	GACCCGTAGG	180
TCAAGTTCAA	CTGGTACGTG	GACGGCGTGG	AGGTGCATAA	TGCCAAGACA	AAGCCGCGGG	240
AGGAGCAGTA	CAACAGCACG	TACCGTGTGG	TCAGCGTCCT	CACCGTCCTG	CACCAGGACT	300
GGCTGAATGG	CAAGGAGTAC	AAGTGCAAGG	TCTCCAACAA	AGCCCTCCCA	ACCCCCATCG	360
AGAAAACCAT	CTCCAAAGCC	AAAGGGCAGC	CCCGAGAACC	ACAGGTGTAC	ACCCTGCCCC	420
CATCCCGGGA	TGAGCTGACC	AAGAACCAGG	TCAGCCTGAC	CTGCCTGGTC	AAAGGCTTCT	480
ATCCAAGCGA	CATCGCCGTG	GAGTGGGAGA	GCAATGGGCA	GCCGGAGAAC	AACTACAAGA	540
CCACGCCTCC	CGTGCTGGAC	TCCGACGGCT	CCTTCTTCCT	CTACAGCAAG	CTCACCGTGG	600
ACAAGAGCAG	GTGGCAGCAG	GGGAACGTCT	TCTCATGCTC	CGTGATGCAT	GAGGCTCTGC	660
ACAACCACTA	CACGCAGAAG	AGCCTCTCCC	TGTCTCCGGG	TAAATGAGTG	CGACGGCCGC	720
GACTCTAGAG	GAT					733

## (2) INFORMATION FOR SEQ ID NO: 2:

## (i) SEQUENCE CHARACTERISTICS:

5

- (A) LENGTH: 5 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

10

Trp Ser Xaa Trp Ser  
1 5

15

## (2) INFORMATION FOR SEQ ID NO: 3:

## (i) SEQUENCE CHARACTERISTICS:

20

- (A) LENGTH: 86 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

25

GCGCCTCGAG ATTTCCTCGA AATCTAGATT TCCCCGAAAT GATTTCCTCG AAATGATTTC

60

CCCGAAATAT CTGCCATCTC AATTAG

86

30

## (2) INFORMATION FOR SEQ ID NO: 4:

35

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

40

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

GCGGCAAGCT TTGTGCAAAG CCTAGGC

27

45

## (2) INFORMATION FOR SEQ ID NO: 5:

50

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 271 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

55

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CTCGAGATT CCCCAGAAATC TAGATTTCCT CGAAATGATT TCCCCGAAAT GATTTCCTCG

60

60



AAATATCTGC CATCTCAATT AGTCAGCAAC CATAGTCCCG CCCCTAACTC CGCCCATCCC 120  
GCCCCCTAACT CCGCCAGIT CCGCCCATTC TCCGCCCCAT GGCTGACTAA TTTTITTTAT 180  
5 TTATGCAGAG GCCGAGGCCG CCTCGGCCTC TGAGCTATTC CAGAAGTAGT GAGGAGGCTT 240  
TTTTGGAGGC CTAGGCTTTT GCAAAAAGCT T 271

10

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:  
15 (A) LENGTH: 32 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

GCGCTCGAGG GATGACAGCG ATAGAACCCC GG 32

25

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:  
30 (A) LENGTH: 31 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GCGAAGCTTC GCGACTCCCC GGATCCGCCT C 31

40

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:  
45 (A) LENGTH: 12 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GGGGACTTTC CC 12

55

(2) INFORMATION FOR SEQ ID NO: 9:

60 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 73 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

GCGGCCTCGA GGGGACTTTC CCGGGGACTT TCCGGGGACT TTCCGGGACT TTCCATCCTG 60

10 CCATCTCAAT TAG 73

15 (2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 256 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

25 CTCGAGGGGA CTTTCCCGG GACTTTCCG GGACTTTCCG GGACTTTCCA TCTGCCATCT 60

CAATTAGTCA GCAACCATAG TCCCGCCCT AACTCCGCC ATCCCGCCC TAACTCCGCC 120

30 CAGTCCGCC CATCTCCGC CCCATGGCTG ACTAATTTT TTTATTTATG CAGAGGCCGA 180

GGCCGCCTCG GCCTCTGAGC TATTCCAGAA GTAGTGAGGA GGCTTTTTTG GAGGCCTAGG 240

CTTTTGCAAA AAGCTT 256

35

(2) INFORMATION FOR SEQ ID NO: 11:

40

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 582 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

GGCAGGAGT AATTCTACC AGAAATTCC AGAGCATTAT GTAGGTAGAA AAAAATGCAA 60

50 GCAAGCTGTT AAAGATCTTG GATCCCATTA TATAGTATGT ATAGCTGAAA TCTGTAATTC 120

AATCACTTTT TCTCTTTTAT CCTCTAACCA AAAAATTGTT TAATTTTGCA TCCCAAATGT 180

55 TTTTAATCTT TGTATATTTT TTAAAAATCC TTTTCTCCTC ATCATTCGCT TTTTGTGGT 240

TGTAATAAGA CTTACTTGCA CTTGAAGAT GAGTTACTCC TTGTCATCTT ACAAATATGT 300

GATATGGTAA TTTTCATAAC AGATGTCAGT TTTGAACCA GAATTGGTGA TTTGTTTATA 360

60 AGAAAAAAC TGGCTTCATT TCTGTGAAAT TGCTCTTGA AAATTTCTTT TTACACGTGT 420

	AAGCCAAC TG AGATACCG TG ATGGTGTTGA TTCTTTTCAA TGATGCTTAC CATCTATTTT	480
5	AGCCACTGAG CCTTTTATTA TTGTCTATT TGTAAGTTT ATTGTCTTA ACTCATTTAA	540
	TAAATATACT GTTTATCTGT TTCTGAAAAA AAAAAAAAAA AA	582
10	(2) INFORMATION FOR SEQ ID NO: 12:	
	(i) SEQUENCE CHARACTERISTICS:	
15	(A) LENGTH: 465 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:	
	GTTTGGGGT GAGGCCGAGC TGCTGCGGG CTTCGTGCGC GGCCAGGACA CAGCTACTCG	60
	CACGGCGGCG GCGCCTGGCT ATGATGTTC TCACCCAGG CGGGCCTCTG CCCTCTACTC	120
25	GTGCCAGGCC CACTTGCCAG GCAGGAGCCC TCCCAAGCC TTCAGGCTG CTCGGAGTCA	180
	CCTGTTGGAA TGGACTAAAA GGACCCTGT GTGGGAACAG GTGCTCCCA AACACCCTGC	240
30	TGCTGGCTGC CAGGCAGGCC CTCTGGAAG GAAGGGCAG GACTCATCAG GACCTCCCTG	300
	GACCCCTGCA GGGCAGGCAG CTTGGGCCG AGCCAAGCA TTGGCTCTG CTGCCCCAA	360
	GGGACAGGA AGCCTCTTGG GCCTCTTCCC TTCTGGACA AGGCCCCCTG CCTTGGCTC	420
35	ACATAAATG TACAGTATTT TCATTAAAG CCTCTTTCAT AAAAA	465
40	(2) INFORMATION FOR SEQ ID NO: 13:	
	(i) SEQUENCE CHARACTERISTICS:	
45	(A) LENGTH: 474 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:	
50	ATGCAATTC TGCTCACAGC CTTCTGTTG GTGCCACTTC TGGCTCTTG TGATGTCCC	60
	ATATCCCTAG GCTTCTCCCC CTCCTAGAAG GGCTTCTGA TAGATTAGAA AATAAGAATG	120
55	AGTGACATTT CCTATGTGCA TATAAGAAG AGCCACAAGA CATGTCTTTT AAATAAAAGG	180
	ACAGTGTCCA TCCTTTTAGC TGCCGAATAG AACCTTGGTC TCATCCTCCT GGAGCTAGGC	240
	CTTTAAACA GCTTCTGTGT TTCTCATTTG TCTCAGTGT TTGCCAGGT TTTATCGGAA	300
60	AGATAATGTT CCGTTTAAAA TATTTCTTAA TGAGGCCGGG CGTGGTGGCT CAGCCTGTA	360

ACCCTAGCAM TTGGGGGCTG AGCGGGTGGA TCACGAGGTC AGGAGATCGA GACCATCCTG 420

GSTAACATGG TGAAACCCCG TCTCTACTAA AAATACAAAA AAAAAAAAAA AAAA 474

5

(2) INFORMATION FOR SEQ ID NO: 14:

10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 314 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

TTATGTTGGG GAGCAAGACC TGATAGCCAG CCTTTACATG GGAGTATAAT TCTGTCCTCC 60

20

ATCTCATAAG CCCAGTACC TGAGCCAGAA TGATTATAAC CAACCACACT GTCTCTTTAT 120

CATGGATGCC TTTAGCAGTA GGTATTTTC ATCATTGCCA TTTGTAGCTC TACAGTGGTT 180

25

TATAGTAATT TCTCATCTTT TAAGTCTCTC CCTCAGTGCC TGTGTGTATC AAATCATTTG 240

CTCTCTCANG CAGTTGAGCT CTGCATCTCT CCYTATGGGG GAGAGCTGTG TTGGAGAGAG 300

AGAATAINAC TTCC 314

30

(2) INFORMATION FOR SEQ ID NO: 15:

35

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 613 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

CTCATATGCG CGTCTGGCTA AAAGTGAACA TGCCATTGAT CAATCTGCTT TTATTATATT 60

45

ATGTTCTCTAA TGGTGGCAAG CAAGACAAGA AGTAGAAAGA AAGATGGTGT AAGCTCAAGA 120

ACCCACTAAA TCTATCCTAT GGCTGGGTT CACCCAGCCT GCTTTGTGGA TTTGTCTCA 180

50

CTATAACAGA GCTCCCAAGG AGACTGCAGA GTCAGCTCCC TTAAGCACTG TAACTAAAGC 240

CTAACTCTTC CGTTCCACCC AACAATGTYC CCAGCTCATC CTCTTTCCCR AAGTCCCCTT 300

TCTGCCCCAG ATGCGAATTG CATTTAACTA ATCCTCAAGT GAAATGTCCA CACAGRATTC 360

55

CATTTTAATT AGCATACCAT AGTTTTGTG CAAATTTGCT TTCAGARGAC TCCCATTGCA 420

GCTGCTCAGA GACGCTAAG GCAGGGCCTC TTGAWGCTTT CCCGATAGCT TTCAGCTGCA 480

60

ATAGCTCTTA GGCAGAATGC CATGAGCGTC CTGCCCCAAT GTATTACTGG GGAACACCTG 540

ATTGGCTAGA AGTTGATCCT CCTGTAACCTT TTCTGAGTTC TTTACATTTA CTCGTGAAAC 600  
CCAAATATGC CAC 613

5

10

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 356 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

20 CCCCCCAT TGAACCTGG GCTGTGAAAG TTTTGCCTG TGTGGGTCGT TCTGTGTGGC 60

GCCTGGTGTG TGGKTCCCAA CTCCTGTTGC AAAGTGGCAG CAGCCAATCA TGAAGCGCCC 120

TTATTTTATG TTGCAGATGA CCAGGTCTCC CCCCCACAGC CTCTGTCTGG TCCCTCATTTG 180

25 GTGAGTGGTC TGCCTGCCCA AGGAGCCTGA TTGGTGGGAA ATGGCATCAT CTAATATGAT 240

GGGAAGGCAT TTGGTCTCTG TTATGTTTAT TACAACATCA TTGCACTCTG GGACTCCAGT 300

30 CCCTGAAAAC GTAATTTGTG GTGTTACCAA AGGACCACAG GGGAAAAAAA AAAAAA 356

35

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 414 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

45 GAAACTANAT CCCGGGGCTT TTAACNGGTA CTGGGAAAT AAGTATTGGG TAATCACTAA 60

GNGGACATTG ACTGCACCAA ACCAAAGCTA TAGAAAGAAA TGATTGACTT TTAAAAATAT 120

ATTCACATTA ACTGTCCTAG GATACTTCTC TTGAGGCTTT GGAAAACTTC TTCCTTGAAA 180

50 TTTCATATC CACTCCAGTT CTGTCACCAA AGATTTTAAT CTTGAGATCG CAATTTCTTC 240

TCTCCAGAA AAAAGTACTA CAACAGGCTC AAGGATATG CTTTGGTGGT CAAGGGATTA 300

55 CACTATGGTT TTCCTTCTGT TCACAATGGT ATTTACAGGA GACCTTGTCA TCAGAGGACG 360

TACTGAACTA TCATTATGAC TTGGATTG ATCAGAGGTT TAAAAAAA AAAA 414

60

## (2) INFORMATION FOR SEQ ID NO: 18:

## (i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 469 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

10 AATCACCATT GCAATACAAA TGATCTGCCT GGTGAATGYT GAGCTGTACC CCACATTCGT 60  
CAGGAACYTC GGAGTGATGG TGTGTTCTTC CCTGTGTGAC ATAGGTGGGA TAATCACCCC 120  
15 CTTCATAGTC TTCAGGCTGA GGGAGGTCTG GCAAGCCTTG CCCCTCATTT TGTTCGCGGT 180  
GTTGGGCGCTG CTGCGCGGG GAGTGACGCT ACTTCTTCCA GAGACCAAGG GGGTCGCTTT 240  
GCCAGAGACC ATGAAGGACG CCGAGAACCT TGGGAGAAAA GCAAAGCCCA AAGAAAACAC 300  
20 GATTACCTT AAGGTCCAAA CCTCAGAAC CTGCGGCACC TGAGAGAGAT GTTTTCGGGC 360  
GATGTCGTGT TGGAGGGATG AAGATGGAGT TATCTCTGC AGAAATTCCT AGACGCCTTC 420  
25 ACTTCTCTGT ATTCTCTCTC ATACTGCCT ACCCCCAAAT TAATATCAG 469

## 30 (2) INFORMATION FOR SEQ ID NO: 19:

## (i) SEQUENCE CHARACTERISTICS:

- 35 (A) LENGTH: 550 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

40 CCCCCCCCC CCCCCACACT TTCAGGAGTC ACCCCCCAGC ATTTGGGGTT GGGTTGGCCC 60  
TACTCCAGCC TGGAGCTCCC TGAGGGAGCC TGCACTCCCT GCTCCCAATC CCGCTACTG 120  
GTGCAGGGAT GCAGCCTGGA GCTGGGTCC TTGTTCTGGG CCTGCTGCTG CCGCCACCCC 180  
45 AGAGCCCCAG CCTGTCTGA ATTGACATCA GTGCTTCCCT GAAGTGCCTC CCCCACCCCT 240  
GGGCATTATC CCAGGAACT TTATGTTTTT TAGAAGCTAA GCAGCTGCTG GGACTCAGGG 300  
50 ACTGGTGAG GTAGGCTGAG TGGCAGCTCA GTCCTAGAAG GTCTCTGAAG ATCTGGACTG 360  
AGGACCTTGC TACTCCCCAA GCCAGAGCCC ATCAGCCAGG CCTGCTGTGA GCCACCTGCC 420  
TGTGGAGTGC TGAGCTCAAC CAAAGGCTGG CAAGCTCTGG GCCTCATTTA AGGGATTCTG 480  
55 ATGAGCCGAT GGGCCCTGGA GGCAGCCCAT TAAAGCATCT GGCTCGTTTT TGGAAAAAAA 540  
AAAAAAAAA 550

60

## (2) INFORMATION FOR SEQ ID NO: 20:

5

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 741 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

	TCTTGAAGAG TGTACAGTAC AGGATTATTA TAATGAAAGT TTATATCAAC AGGGTTTCGT	60
15	TGGCTCTGCA TATATTATAA GCAAAAGAGA TTGGTAAAGT GCCACAGTAT TCCAGATAAC	120
	TTTTCAGTTG CGGCCTTTCT TCTCGTTCTT TAATTTGAAA CCTAGATACA TGCAGTAAAA	180
20	ACTAGGAGAA TGACTTTTAC CCTTGGGGAC AGCCAAGTTT TGTGATAAA CCTATTTCTT	240
	AGCATGCCTT CAGGAAGTTG TGCCAGACCC TAGATTGTGA AGGACCCACT GTTCTTCTGT	300
	TGTACGAGCT CCTGAACCA TTGTTTCAGAG GACCAATGTC ACATCGCTTC ATGGGCATGG	360
25	NCCATGGGAG CATCTGGGTG ATAYCTGTCT ACAGTATTGG CTCTTCTGCG AGGCTGATAC	420
	ACAAGGCCTC TCTTCCACAT GATCATTGTC AAACCTCCCC CAGCCCCTAC CATCCAATGT	480
30	GGAAGGAAAA CAAGAACTGC CTGAAGAAGA GTCCAAGCTA CAGATACACA GCGTGTGCAT	540
	TGCGGCTGTC ACCTTCCTCC TCCCCTTCT GTATCCTCAG AGATGCTGCG TGGATGTTTC	600
	CTTAACCTCA GCTGACTTCC CTGTGAATGT CTAATGCTAG TTCAGGGCCT CCAGGCATTG	660
35	ATTGTGACAG TGGTAACTCC CAATGAGGCT TCTGTTATCA TTTGGTGTGC TTTTCTGTC	720
	ATTAAAAGAA ATGATTTTCC C	741

40

## (2) INFORMATION FOR SEQ ID NO: 21:

45

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 991 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

50

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

	GGCAGGAGTC TCCCCTGGGG AAGTTTTTCT TTTTCAGGAG GGAGGAGGGC TTTCCCAGGT	60
55	AATGTGTCTA GAGTGTGGG CAGAAAATCT GGGACCACAC CACACCAGTT CTCTCCTTAA	120
	TCCACGTCAT TTGCCTTCTA TCCCAGCTAT GTTTCAGTG TCCTCTGGGT GTTCCAAGA	180
	GCAACAAGAA ATGAATAAAT CTCTGGTGAG TTGTTTATTT GTTCTTCACT TTGTTTACA	240
60	CTGTATTTTC TGAGTTTATG GGTGTCTGTG AATTAAAAAG GAAAAGTAGA AATAAGTAAA	300

ACTCAGGTTG AAGGAAATAT ACATAAATAA GATAAAGCTG ACCTGTAGAT ATAGCAGGTT 360  
 ATAAAGCTTA GAGTTGTCTA AGTTGAGTGC AAATTTTCCT CTGATCTTTC TGATGCCGAA 420  
 5 CAAAAAGCA GTCATGTTTG TTATGTGATT GGAATGGAAC CCGAGAAGAG AGCATGCTGT 480  
 GTTCTTGTGG GACAGGAAAG CTTGCGTGCA CCAAGTCTGA ACCACCACCT TCATGGTGAC 540  
 10 ATAGATTATG TGCTGGAACA TATTTACAC CCGCCTGGCA GTAAACACTT GTAGTGTGT 600  
 GCAGTGAAA CGTCATCTT CGCTAAAGC ACGGCGTGT GTGCAGCGGA AATGGTCATC 660  
 TGCTGCTAAA ACACAGCTTC CATCGTAATG TATGCTCCTT ACTCAAAGAG TGTGGTCCCA 720  
 15 AACAGCCTTT GGGAGGTCTT CCTTGATTCA TGGATGAAAC CTGGAACATC TTGAGGACTG 780  
 AGTTAACCAT AGGTCCTTAA ATAACCTCTC ACACGTTTTT CTTAGTTTAT CTCTACATGC 840  
 20 AGGGTGTGCA GCAGCCTGTT CAAAGTCATA TTTTCTGGGA AATATTTCCA GTGTTTATTT 900  
 GCACCTTAGC CCACTCTGTG TAGCCTTATT TCTTCTAAAC TCACCATTAA TCTGAATAAT 960  
 AGTCAAATTT AGGGGGACTG TATTTGCCTT A 991  
 25

(2) INFORMATION FOR SEQ ID NO: 22:  
 30

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 653 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 35 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

CCACGCGTCC GGAATTCCTT TGAGGATCTT GGGCTATCTT TGACAGGGGA TTCTTGCAAG 60  
 40 TTGATGCTTT CTACAAGTGA ATATAGTCAG TCCCAAAGA TGGAGAGCTT GAGTTCTCAC 120  
 AGAATTGATG AAGATGGAGA AAACACACAG ATTGAGGATA CGGAACCCAT GTCTCCAGTT 180  
 45 CTCAATTCTA AATTTGTTC TGCTGAAAT GATAGTATCC TGATGAATCC AGCACAGGAT 240  
 GGTGAAGTAC AACTGAGTCA GAATGATGAC AAAACAAAGG GAGATGATAC AGACACCAGG 300  
 GATGACATTA GTATTTTAGC CACTGTTGTC AAGGGCAGAG AAGAAACGGT AGCAGAAGAA 360  
 50 GTTTGTATTG ATCTCACTTG TGATTCGGGG AGTCAGGCAG TTCCGTCACC AGCTACTCGA 420  
 TCTGAGGCAC TTTCTAGTGT GTTAGATCAG GAGGAAGCTA TGGAAATTAA AGAACACCAT 480  
 55 CCAGAGGAGG GGTCTTCAGG GTCTGAGGTG GAAGAAATCC CTGAGACACC TTGTGAAAGT 540  
 CAAGGAGAGG AACTCAAAGA AGAAATATG GAGAGTGTTC CGTTGCACCT TTCTCTGACT 600  
 GAAACTCAGT CCCAAGGGTT GTGTCTTCGG AGGCATCCAA AAAAAAAAAA AAA 653  
 60



## (2) INFORMATION FOR SEQ ID NO: 23:

5

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1486 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

10

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

15	GGCAGGCTGA CGACCTGCAA GCCACAGTGG CTGCCCTGTG CGTGCTGCGA GGTGGGGGAC	60
	CCTGGGCAGG AAGCTGGCTG AGCCCCAAGA CCCCAGGGGC CATGGGCGGG GATCTGGTGC	120
	TTGGCCTGGG GGCCTTGAGA CGCGAAAGC GCTTGCTGGA GCAGGAGAAG TCTCTRGCCG	180
20	GCTGGGCACT GGTGCTGGCA SGARCTGGCA TTGGACTCAT GTGCTGCAT GCAGAGATGC	240
	TGTGGTTCCG GGGGTGCTCG GCTGTCAATG CCACTGGGCA CCTTTCAGAC ACACTTTGGC	300
	TGATCCCCAT CACATTCCTG ACCATCGGCT ATGGTGACGT GGTGCCGGGC ACCATGTGGG	360
25	GCAAGATCGT YTGCTGTGC ACTGGAGTCA TGGGTGTCTG CTGCACAGCC CTGCTGGTGG	420
	CCGTGGTGGC CCGGAAGCTG GAGTTTAACA AGGCAGAGAA GCACGTGCAC AACTTCATGA	480
30	TGGATATCCA GTATACAAA GAGATGAAGG AGTCCGCTGC CCGAGTGCTA CAAGAAGCCT	540
	GGATGTTCTA CAAACATACT CGCAGGAAGG AGTCTCATGC TGCCCGCANG CATCAGCGCA	600
	ANCTGCTGGC CGCCATCAAC GCGTTCGGCC AGGTGCGGCT GAAACACCGG AAGCTCCGGG	660
35	AACAAGTGAA CTCCATGGTG GACATCTCCA AGATGCACAT GATCCTGTAT GACCTGCAGC	720
	AGATCTGAG CAGCTCACAC CGGGCCCTGG AGAAACAGAT TGACACGCTG GCGGGGAAGC	780
40	TGGATGCCCT GACTGAGCTG CTTAGCACTG CCCTGGGGCC GAGGCAGCTT CCAGAACCCA	840
	GCCAGCAGTC CAAGTAGCTG GACCCACGAG GAGGAACCAG GCTACTTTCC CCAGTACTGA	900
	GGTGGTGGAC ATCGTCTCTG CCACTCCTGA CCCAGCCCTG AACAAAGCAC CTCAAGTGCA	960
45	AGGACCAAAG GGGGCCCTGG CTTGGAGTGG GTTGGCTTGC TGATGGCTGC TGGAGGGGAC	1020
	GCTGGCTAAA GTGGGKAGGC CTTGGCCAC CTGAGGCCCC AGGTGGGAAC ATGGTCACCC	1080
50	CCACTCTGCA TACCTTCATC AAAAACACTC TCACTATGCT GCTATGGACG ACCTCCAGCT	1140
	CTCAGTTACA AGTGCAGGCG ACTGGAGGCA GGA CTCTCTGG GTCCCTGGGA AAGAGGGTAC	1200
	TAGGGGCCCC GATCCAGGAT TCTGGGAGGC TTCAGTTACC GCTGGCCGAG CTGAAGAACT	1260
55	GGGTATGAGG CTGGGCGGG GCTGGAGGTG GCGCCCCCTG GTGGGACAAC AAAGAGGACA	1320
	CCATTTTCC AGAGCTGCAG AGAGCACCTG GTGGGGAGGA AGAAGTGTA CTCACCAGCC	1380
60	TCTGCTCTTA TCTTTGTAAT AAATGTTAAA GCCAGAAAAA AATAAAAAAA AAAAAAAAAA	1440

AACTCGAGGG GGGCCCRKAC CCAATCWCC TATAGTAKAC GTANN

1486

5

(2) INFORMATION FOR SEQ ID NO: 24:

10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2323 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

CTTCGCGGTT TCTCCTGCCA GGGGAGGTCC CGGCTTCCCG TGGAGGCTCC GGACCAAGCC	60
CCTTCAGCTT CTCCCTCCGG ATCGATGTGC TGCCGCGGCC GCCGCCGCCG TCCCGCGTCC	120
TTGGTCTCT GCTCCCGGA CCCGGCTCCG CGCAGCCAGC CAGCATGTGC GGGATCAAGA	180
AGCAAAAGAC GGAGAACCAG CAGAAATCCA CCAATGTAGT CTATCAGGCC CACCATGTGA	240
GCAGGAATAA GAGAGGGCAA GTGGTTGGAA CAAGGGGTGG GTCCGAGGA TGTACCGTGT	300
GGCTAACAGG TCTCTCTGGT GCTGGGAAAA ACAACGATAA GTTTTGCCCT GGAGGAGTAC	360
TTGTCTCCCA TGCCATCCCT GTTAATTCCT GGATGGGAC AATGTCCGTC ATGGCCTTAA	420
CAGAATCCCC CAGATGGCTT CATGGCCCC AAAGCATGGA AGGTCCTGAC AGATTATTAC	480
AGGTCCTGAC AGAAGAACTA AGCCTTTGGT CCAGAGTTTC TTTCTGAAGT GCTCTTTGAT	540
TACCTTTTCT ATTTTATGA TTAGATGCTT TGTATTAAAT TGCTTCTCAA TGATGCATTT	600
TAATCTTTTA TAATGAAGTA AAAGTTGTGT CTATAATTAA AAAAATATAT ATATATATAC	660
ACACACACAT ATACATACAA AGTCAAATG AAGACCAAAT CTTAGCAGGT AAAAGCAATA	720
TTCTTATACA TTTTATAATA AAATTAGCTC TATGTATTTT CTAATGCACC TGAGCAGGCA	780
GGTCCAGAT TTTTAAAGC TTTGTTTGAC CATGTGTCTA GTTACTTGCT GAAAAGTGAA	840
TATATTTTCC AGCATGTCTT GACAACTGT ACTCTTCAA TGTCATTTAT CAGTTGTAAA	900
ATATATCAGA TGTGTCTCT TCTGTACAAT TGACAAAAA AAAAATTTTT TTTTCTCACT	960
CTAAAAGAGG TGTGGCTCAC ATCAAGATTC TTCCTGATAT TTTACCTCAT GCTGTACAAA	1020
GCCTTAATGT TGTAAATATA TCTTACGTGT TGAAGACCTG ACTGGAGAAA CAAAATGTGC	1080
AATAACGTGA ATTTTATCTT AGAGATCTGT GCAGCCTATT TCTGTACAAA AAGTTATATT	1140
GTCTAATAAG AGAAGTCTTA ATGCCCTCTG TGAATAATGT AACTCCAGTT ACACGGTGAC	1200
TTTTAATAGC ATACAGTGAT TTGATGAAAG GACGTCAAAC AATGTGGCGA TGTCGTGGAA	1260
AGTTATCTTT CCCGCTCTTT GCTGTGGTCA TTGTGTCTTG CAGAAAGGAT GGCCTGATG	1320

60

CAGCAGCAGC GCCAGCTGTA ATAAAAAATA ATTCACACTA TCAGACTAGC AAGGCACTAG 1380  
 AACTGGAAAA GACCACAGAA AACAAAGAAT CCAACCCCTT CATCTTACAG GTGAACAAAC 1440  
 5 TGTGATGATG CACATGTATG TGTITTGTAA GCTGTGAGCA CCGTAACAAA ATGTAAATTT 1500  
 GCCATTATTA GGAAGTGCTG GTGGCAGTGA AGAAGCACCC AGGCCACTTG ACTCCAGTC 1560  
 10 TGGTGCCCTG TCTACACCAG ACAACACAGG AGCTGGGTCA GATTCCCTC AGCTGCTTAA 1620  
 CAAAGTTCTT CGAACAGAAA GTGCTTACAA AGCTGCCTTC TCGGATACTG AAAGGTCGAG 1680  
 TTTTCTGAAC TGCAGTATT TTATTGCAGT TGAAAAAATA AAAAGCTAT TCCAAAGATT 1740  
 15 TCAAGCTGTT CTGAGACATC TTCTGATGGC TTTACTTCCT GAGAGGCAAT GTTTTACTT 1800  
 TATGCATAAT TCATTGTTGC CAAGGAATAA AGTGAAGAAA CAGCACCTTT TAATATATAG 1860  
 GTCTCTCTGG AAGAGACCTA AATTAGAAAG AGAAACTGT GACAATTTTC ATATTCTCAT 1920  
 20 TCTTAAAAAA CACTAATCTT AACTAACAAA AGTTCTTTTG AGAATAAGTT ACACACAATG 1980  
 GCCACAGCAG TTTGTCTTTA ATAGTATAGT GCCTATACTC ATGTAATCGG TTACTCACTA 2040  
 25 CTGCCTTTAA AAAAAAAAC CAGCATATTT ATTGAAAACA TGAGACAGGA TTATAGTGCC 2100  
 TTAACCGATA TATTTTGTGA CTAAAAAAT ACATTTAAAA CTGCTCTTCT GCTCTAGTAC 2160  
 CATGCTTAGT GCAAATGATT ATTTCTATGT ACAACTGATG CTGTCTCTTA TTITAATAAA 2220  
 30 TTTATCAGAG TGAAAAAATA AAAAAAATA AAAAAAATA AAAAAAATA AAAAAAATA 2280  
 AAAAAAATA AAAAAAATA AAAAAAATA AAAAAAATA AAA 2323

35

(2) INFORMATION FOR SEQ ID NO: 25:

40

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 683 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

GGCACGAGCC TGTGTGGTCA TGTTCCTCGT GGTGCAGTAC CTGACATGAG CCAGCCACGC 60  
 50 TCAGTGCTG AACAGCATTC CCACAGCCTG CAAGTGTGTG TGTGTGTGAA AGAGAGAGGG 120  
 GGGCCAGAG CCGCTTTTG AAATGTTTGC CTGTCTGAAC TGTGAAGACA CTGGGAGTG 180  
 ATTGTGGTCT AATTTCACAC CTGCTCTGTT TTCTGTGACA TCTGGAGGG GAGCTAGTGC 240  
 55 CACACCATGC GCGGTGCTTA GAAATGAAAA AGTCCCGGGT CTGTCTCTCT CACTCTCGCT 300  
 CTCATGGGGG AGGGAAAGAA TGGCTTTGGT GGCTTTGTTC ACACAGCTGA TCGGTGCTGG 360  
 60 GAAGGTGTCC ACAGTGAGCC TGTGTGCAGG ACTGTCCACA CGGTTCACAC TTGTCACCAT 420

5 CAGGCCTTTC TGGTCTGAT AGGGTGGAGC AAAAGTGGAA AGGAAAGGAA AGAGGCTTTT 480  
CTCACAGCCA TTATATTAAA TAGTAGGTCG ATTCACATCT CGTGCTCCTG GCCACCTTCC 540  
CCTGTGCCTC AGTGACATGT AGATGACTGA CTGCCAATAC TTGTCACCAT TCCCTGGAAG 600  
CAGCTACCTA GGGGAAACAA GATGTAGTGC TATTGCCGAT AACAAGTAAG ATTTTCCACA 660  
10 CTAAAAAAAA AAAAAAAAAA AAA 683

15 (2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2036 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

25 CTGAGAAAGG AAAGCATTCG GATCTGCTGC AAAACACAT ATATCCATAA AGACTCATGT 60  
TATTCAGAAA ACAGATTGTG AACACAATCA CATTCGCATG AATCCTTTAA AAGGAAGAAG 120  
ACCTTAAAGT ATCTGCAAAT CTGAATTTCT ATTTATTCCT TCACTGAATA TAGAAACAAT 180  
30 GGTATATCTGA TTATTAGAGA TATTATTTTG GATATGTTAC TTATTAACCT GCTATGGCTG 240  
GTAACCATGA TAAAGTCTGT TATTAATAAC AACATAATC TTTTTTTAAA GAAGAAAAGC 300  
35 TTATTTTTC A TTGACAGTGT ATAGATTTAT CTA CTAGTGT GTGTTTGTCT ATTAGTGTGT 360  
TAATTTTMTT TTTAAGTTGA GTGTTTGATA AATTTTAAGA CCTGTCCCC ACCTGTGTTT 420  
GAGTCTGTG TTGACTACAG GTATATAGCY CAWTTTAAAA ATCCTAAAGC AAAAGAATTT 480  
40 TATTTATAAA AGAATCMAMC MGTTCATGC ATGAGGCTGT GAAGTCAGAT ATTTAGTAAT 540  
AAAAGCAGCA GTGCCTTTTT TTGTATTTAC CCATGACCC CCACCAAATG CAACTGTMTT 600  
45 ATATTAAGAA AATAGTAACA ATTTTAAAAT CTCAGAGTAA AATCTATTTT ACTACATGCT 660  
TTTCCCCCTT TGTCTGATT TAAGCAGTGT GACTTTGGCA TCTCTACATT GTCCTAGGGA 720  
CAGTGGTGT CTACAATATT ATCATGTATG ATGTTTATT GTGCTTTTT ATTCATAGTG 780  
50 GCTTCTTACC AGAAACAGTA GGAAGAAACA CATGAACTGT GTACAAGACA TGAACATGT 840  
CTGCTGATAT GTTGTTTTTT CACATGCTTT TGAGTTTCA CTTTTTAAAC GAGAGCCAGC 900  
55 AAGCAAAATA GATGTGGCTG GGTCTGCCTG TCCGGGCGGC TTTTGCACC GAGCTCTCAA 960  
ATCCTGTGTA TTGAGGTTT CTTTTTGGTA CTCAGGATTG GAGCTACAGC TGGGCCCCC 1020  
TCTCTCCCAT TCGTTTGAAG AGACACTGAG GGAAACAAGG GTTCTTTTTG AGGTGTCCTT 1080  
60

	GGCTGCCTTT TACGGGATGG GAGCCTTCTC CGGATCTTTT GTTCTTCTGC ACCTCTTGTA	1140
	GCTACTGCCG GTGCAAGGTT GTAGATGTTA TTCCCCAGGA GCCTGGGCTK GGGGGCTGAG	1200
5	CTGGGCTGAA TGCAAAAGCA TGCAACCAGA AGGCGGGCAA GGGGAGGAAA AGCAGGCCTG	1260
	GCCTCATTGG TCCCCTGGAG ATGTCTGTAG CAGTCAGCTC CAGCTTGGGC CTGGGGAAGC	1320
10	AGCCTGACCA AGGCGCTCAG GTGTGCCTGT TACAAGAAGA ACCTGCAGAA GGATAATTTG	1380
	CACATGGAGC TGTGATAACA CTAATGTTGA TTTTTTTTTT TTTTACAAGT CATCAGRGAT	1440
	GTTTGCAAAG TGAGTTTAT TTTTTGTAA TTCCTTTATC TTTACTTAAA GGTGAATGTG	1500
15	TATTCCTCTG GGAGGAATAG GAAGAAAACA GGAATGTTAA TAATGTCGAA CAGAAAACIT	1560
	CCTCCCTTAT TAATATATAA TCYTCATGTA TTTATGCCNT AATGTAAGCT GACTTTTAAA	1620
	AAGCTTTCTT TTGTTGCATG CCCTGTGCAG GCATCTGTAT TGTACATGCA TGCCTTTCTG	1680
20	CCTGTTTTCC TGTATAAAGT TAGTGAACAA AGAAATATTT TTGCCCTAGT TCATGTTGCC	1740
	AAGCAATGCA TATTTTTTAA ATTTGTCATA TATGGAAAGA GCATGTTTGT TACATGTAAA	1800
25	AGCTTTACTG ATATACAGAT ATACTAATGT TTGAAGATGC TGTCTTTGC AAGTGTACAG	1860
	TTTTCAAATG TTGTTACCAG TGAAACACCC TTGTGGTTTA AACTTGCTAC AATGTATTTA	1920
	TTATTCATTT CCTCCCATGT AACTAAGAAT CATGGCTATA TTTCATATCA ACGTTATATT	1980
30	GAAAGTGAAG GGAAATGATT AATACAAGGT TTTGTAACAA AAAAAANAA ANNAAA	2036

35

(2) INFORMATION FOR SEQ ID NO: 27:

## (i) SEQUENCE CHARACTERISTICS:

- 40 (A) LENGTH: 717 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

45

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

	GGCAGGAGAT AACATAGGCA CAATAATACT GTATGTCTAC TTCTAGGATT ATAAGGAATT	60
	AACATTGAGA TGACATTTCC ATTTGAGAAG AAAATAGTTG CTTTCAGTGC CTTTTATTTG	120
50	ATTCCTGGAG AGAGCAGACT CGCACCAACA TTCAACCCCA GCGCTGATAT GACAGTAATC	180
	CTCAGAGGCA GAGCCAGCA CAAAACAGCA ATGCTAGAAA GTTACAATTG GAAAGTTTCC	240
55	TGCCAGCTTC GGAATGACA CTGCAAGCT GATGCCAGAA ACTGCCAGAG TAATTCCTCT	300
	CATTACTGCT CTACCCACCC ACTTTCAGCT CCCCAAATTA ACTAGTGCAG TTGACTAATC	360
	CTCTTACCT TTATCATTTA GGTGAGGCAT TGCACAAAA CTCTCGACTT TGCCATATAA	420
60	GGGCTGTGGT TCTCTGTGGT CCTGGATAAG AGGCATCACC ATTATCTGGA AACATGCAGT	480

AAATGCAGAT TCTTCATCTT CTCCCCAGAC CTCCTGAGTT AGAAATTCAC AAGTTCTCCA 540  
GGTGATCTCA TACATGCTAA AGTTTGAGAA CCATTGAGTA AAGTTAATGC ATTAAGAAGA 600  
5 GATTAGATAG GGATGGTGGC GTATCTTCCT ACAGTTTCCC TGTTAACAAG AAAGTCAGAG 660  
GTCAGTTGAT CAGACATTAG ATTATTTATT GCTAAACTA AAAAAAATTA AAAAAA 717  
10

(2) INFORMATION FOR SEQ ID NO: 28:

15 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 495 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
20 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

GAATTCGGCA CGAGCAGCAT CCTAATTTTA GTTTGGAGAT GCATTCTAAA GGATCTTCTC 60  
25 TATTGCTTTT TCTCCACAA TTAATCTTGA TTCTGCCTGT CTGTGCACAT TTGCATGAGG 120  
AACTGAACTG TTGTTTTCAT AGGTAAATGA GAGACTGAGT TTTTTCATTT CTGAAGAGAA 180  
AGGGCATTTG CTCCTACAAG CTGAAAGGCA CCCCTGGGTG GCTGGGGCCC TCGTGGGAGT 240  
30 TTCTGGGGGA TTGACCCCTA CAACATGCAG TGGCCCTACA GAAAAACCTG CAACTAAAAA 300  
TTATTTTTTA AAAAGGCTCC TCCAGGAAAT GCATATAAGG GCTAATCACC CAGTATTTTG 360  
35 ARGCTTCGAA GARGTAATAR AMCCCTGGAG AGAGAACTG AGACATGTAA GAGGGTGGGA 420  
ATGACTCAGT GGTGGCACAC TATGGAGTCC TGCCACAAG TAGCACACAT CAACCCACTA 480  
CACAGAAATC CTAGG 495  
40

(2) INFORMATION FOR SEQ ID NO: 29:

45 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 556 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
50 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

AGCTTAACGT CATGATTCAT TAGGGGAATG CAAGGCAAAA CCATGATGAG AATGCCCTTA 60  
55 GACACCTCTT AGAAGAGCTG CTAGAAAGGC AGACAGCACC AAGCGCTTAA ATGAGATGGG 120  
GGCACTGGTG CTTCTTCTGT GCCTACTGGT AGGGGTGCAG CAGAGTGGTT CAGTCTGGGA 180  
60 CAGTTAGCTG GACATCACGT GGACCCAACA CACGCATTTC CTGGGTACT TACCAAGGAG 240

AATAGAAAGC AGGCAGATCT TTACAGCAGC TCTTACCTGW TTGCAAAACA ATGGAAATGC 300  
 CCACATGTCC ACAACAAGT KTGTGGTCTG CCTGTGCCAT GAAGCACAGT GTGGCTGAGC 360  
 5 GTCAAGAGTC CCCACACTCA AAGGAGGCAG CAGATACAGG GCTGCACACT GTGTGATTCC 420  
 ACACATGTGA CATCTCTGGAC ACGGACATGC TGGATGGCAA AACGAGCATC GGGCTGAGAG 480  
 10 GACTGCTGAG AAGGGGAACG GGGCTGCTGG GATGTGGGTT GATTGTAGCA GTAGCTCATG 540  
 GAGATGTGAC CTCAA 556

15

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:  
 20 (A) LENGTH: 434 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

CTAATGGTG ACTGTGGCTT TGTGAGACA GGCCCCAAAT GGTAGGTGTG AACACAACAT 60  
 GCACAGAATG AGGAGACATG CAGAGTGTCT AAATACTGTC CTGGACAGAT GTGTTACATG 120  
 30 ACTTCTTTT CAGCTTATTT CTGTGGCTG CCTTTGAAGA TAGAGCTTTG TTGATATTTA 180  
 CATTAAACCA AATTGTATAA YTATGTTCCA TTCTGACATG TTATTTAGCA AARGAAAAAR 240  
 35 GAGTAATCT ACATCAGCAT CTTTAGTGCA TGCTAAAAGA TTAAAAATGT CTTTGGGGA 300  
 ACATGTTTGT TATACATAAA TGTTAGATA GAAATATTTA TAGAATNCTC TATGTGAGTA 360  
 TTATCTCCC TATGTATATT TATATCTAGA TGTGTCAATC TTGTATTGA TATGAAATGC 420  
 40 TATGAATAGT GAGA 434

45

(2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:  
 50 (A) LENGTH: 715 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

CCACGGTCC GATCTCACAG CTCGACACT ATTGCGAGCC ATACACAACC TGGTGTGAG 60  
 AAACGTACTC CCAAACTAAG CCCAAGATGC AAAGTTTGGT TCAATGGGGG TTAGACAGCT 120  
 60 ATGACTATCT CAAAATGCA CCTCTGGAT TTTTCCGAG ACTTGGTGTG ATTGGTTTGT 180

CTGGCCTTAT TGGACTCCTT TTGGCTAGAG GTTCAAAAAT AAAGAAGCTA GTGTATCCGC 240  
 CTGGTTTCAT GGGATTAGCT GCCTCCCTCT ATTATCCACA ACAAGCCATC GTGTMTGCC 300  
 5 AGGTCAGTGG GGAGAGATTA TATGACTGGG GTTTACGAGG ATATATAGTC ATAGAAGATT 360  
 TGTGGAAGGA GAACTTTCAA AAGCCAGGAA ATGTGAAGAA TTCACCTGGA ACTAAGTAGA 420  
 10 AAATCCATG CTCTGCCATC TTAATCAGTT ATAGGTAAAC ATTGGAAGTC CATAGAATAA 480  
 ATCAGTATTT CTACAGAAAA ATGGCATAGA AGTCAGTATT GAATGTATTA AATTGGCTTT 540  
 15 CTCTTCAGG AAAAAGTAGA CCAGACCTCT GTTATCTTCT GTGAAATCAT CCTACAAGCA 600  
 AACTAACCTG GAATCCCTTC ACCTAGAGAT AATGTACAAG CCTTAGAACT CCTCATTCTC 660  
 ATGTGCTAT TTATGTACCT AATTAAAACC CAAGTTAAAA AAAAAAAAAA AAAAA 715  
 20

## (2) INFORMATION FOR SEQ ID NO: 32:

- 25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 486 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 30 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

GAGCCAGTGC CGGCGAAAGG GGACCTTCCT CTACTTCTG CCACAGACCC TGTCCCCACA 60  
 35 CACTTCTGCG CCCTGCTCTG CTGGGAGGCC ACTTCCTCCC CCAGTGCTGG ATTCCACCCC 120  
 CAGCTACCCC TCAAACATGG CCCCCTCTCT CCTCTGCTT GCCCTCTCT GCTCCCTGGA 180  
 40 GGCTGTCTG TCCTCCCCTC TTGAAAAGCA ATGCCAGCTT CCTGGGATCT TCTGCCAACT 240  
 CCAGCTACCA TGCCCTTTGC TCCTGTCAGC TCAGCTCCTC AAGGGAATTG TCTAMCCTCG 300  
 GTGTCTGCT TCCCTCCCTC AACCTCCTCA CCCTGCTCCA AGCTGGCATC TGCCCTCCA 360  
 45 CTGCACAGAA CGGNTCCCCC ACCACCTGCC TTTACAGGGA GGAAGCAGCA ACATGGAAGA 420  
 ANCGAATAT AGGGGCTACA ANGATGCTCA GCTCTGATCC CGAAGGCAAA AAGNATCTTT 480  
 50 GGGCAC 486

## (2) INFORMATION FOR SEQ ID NO: 33:

- 55 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 725 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 60 (D) TOPOLOGY: linear



## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

5 GTTCCTCTGG TAATAATTAG GTTATTCCTCA GAAGCACAGT GTCATTCTTT AAATAAAAGC 60  
 TTTCCTGTTT AAAGCTTTTC AAAGGAGCAG ACCACCTTGA AGATTCCCCC TAGGGTIGAT 120  
 ATGTGTCTAA TTCATTTTAT AAAAATTATT CTGTCTTCA TTTTAAAGCT TTGGCTATAT 180  
 10 AGTCAGAAAT GTCCTAAATA ACAAACTATT TTGTATTTAA TTTAGGGAAG ACTAAAGGGA 240  
 AGAAAAATGA AAATCAGTC TTTATGTAAG CTCCAAGGAT ATTAGGGCTT AAAGGGCTTT 300  
 TCTAGTTTTA TGAGAAATTG TACTACTGAT TTTTATATAT TCCTGTTTTT GATGAACAGA 360  
 15 TCTCTGGGGA AATTGTTGAG TTACAATGGC ATTTCACTGT GATCCCTCTC AAGCTCAGAT 420  
 CAGTTCTATA ACCCAATGAC AACCTGTCTC TTTGGTTTAC TGTCTGTGA AATGTCAGCT 480  
 20 CAAGTTTCCC AGAAGTCGTG TGTATTATGAT GAGTCAGAGT GCTTTTCCTC GGTGGGACAG 540  
 TTGCTGGCCC TCTTAATTTT GGTGTATGTG CTCCAAGTA TCTAAACCTC CAGTCTGATC 600  
 TGTATATGCT ATCCTAACTG TTAATTGTAT TATTGATTAT GTTGATTATC TTGCTTGAAG 660  
 25 GTTCATACTT TTCAATTGA TAGAAATAAA GTTTTTTCT GCTTATAAAA AAAAAAAAAA 720  
 AAAAAA 725  
 30

## (2) INFORMATION FOR SEQ ID NO: 34:

35 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 437 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear  
 40

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

CACACAGCAT GCTGCCCTCA GACGTGTCCA TCCTGTACCA CATGAAAACG CTGCTGCTCC 60  
 45 TGCAAGATAC TGAGAGATTG AAGCATGCTC TGGAAATGTT CCCAGAACAT TGCACGATGC 120  
 CTCCTGCTTT TATTGGCTCT TGTGAAATC AAATTGGAAG ATCTTCAGTC CCAGCTGCAC 180  
 50 CCAACGTGGA AAAGTATTCC AGGTCCATCC CCAAGGAACC AACACCGATG ACATGGACTC 240  
 AGGAATCTTA TAACCTACGT GGACTCTTTC CATCCGTACA TTGTCGTGCA CATGCCACTC 300  
 ATCACCTGGC GTGCCCAGAT CCTCGCARGG CAACACCTG TGATAATTCC AGGTGATTCT 360  
 55 CTACATCTGC AGCTTGAGGT TAGCCTCATA TCACATTACA TTCTCACTAN AAACNAAAAA 420  
 AAAAAAAAAA AACTCNA 437  
 60

## (2) INFORMATION FOR SEQ ID NO: 35:

## (i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 943 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## 10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

GGCACGAGCT GGAACAGAGA CTAAATCCCA CGAAACTGAC ATTGTTAAAC AACTAAAAAC 60  
AGAAGTACTT ACCTCTTGAA GATTTAATAT ATAATGGTTG ACATGATACA TGTACATGAT 120  
15 GAATGACCAG ATGCTTATGG TCTACATTTT CCTTTATCCT GTTAGTATTA CCTTCCTTAA 180  
TCTTTGTICA TTAACATGCT AATTCCTCTT CAGTGTATTAT TTTCTAGTGA CAGAATGCTA 240  
20 ACATTTCTTA CACCTGGCA GAAGGGAGAG AAATGTGTTT TGGGGTGGGT AACTAAATTT 300  
TTGAGTGAAA TATCATAAGA TGANAATGGA AANAAGGAGA CACAAANAGT TATNACAAAA 360  
AAACAATGGT TTTTPTAGCC ATTTGACTGG CTCTTTAAAT AGTCTACAAG ACATTCACGT 420  
25 TTAACATCAC TTTTAGTGAA ATAAATGTG CCATACTAGT ATGTGCTTCA AAAGGGCAAA 480  
TGTGCTTTAG TGCCCTAAGG CTAAATTTTG GTCATTTGAC ATCAGAGATG TTGTAAGTAT 540  
30 TGCACTTAAT ACGCACCTAT TTNTCAATAG TGTATTTTTT TGGNTAGCAT TTTTTTTACC 600  
ACTATNTTGT TGATAGCTTT TTGTTCTNTN AGGTIGNAAN ATGACAGTGC TNATNTCAAA 660  
CAGATTACCC ATNTGCAGAA CTAAGGAAG CNATTTATGT ATGAAAGNAA TTNTTGAATT 720  
35 NGTCATNTIC AACNNTGNA TTAAAGCTTA GACTAAATAG TAATATATNG TGGGNAGGAT 780  
TTTGGTTTGG TGATATTTNT GTGNATTAAG GNATAGATGT TAACNNTTAT TTTGTAGNAA 840  
40 AGTGANTTGT ATGTGGTTAA TTATAAATAA AACTGGTACC AGGNAAAAAA AAAAAAAAN 900  
NAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA AAA 943

45

## (2) INFORMATION FOR SEQ ID NO: 36:

## (i) SEQUENCE CHARACTERISTICS:

- 50 (A) LENGTH: 604 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## 55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

GGCAGAGAA ATCTTCATGC TGTAGTCACT CCAGACCATG GAGTGGCTTT CCAGCTGAAT 60  
60 GAATCCTATG TCTCGCGTGC AGGTGGTTGG TTTCAATGT TCTTGCTAAT TTTTITCTA 120

TTGGATCTTG GGAGTTTCTT TTGTTTGCTC CTGTGTTTGC CCAGCTTTAA TAAAACCAGG 180  
 CGCAACAAAA AACCATAGCA TTCTGAACAA TAGGGGGCCC ACATTGGACC CAGTATGTCA 240  
 5 CTTTAATGGA CTTCAAGAAA AAATCTGAAT GGGAAAAATG AACTAGGAA TGTATACTCC 300  
 ACACATTTTA TGCCATATAA TGGTGTGTTT TCTTAATTTT GTTCTTGTG GCGAAATGTG 360  
 GCTTTCAAAT TAAAATGACC TTTTCTTCTT TGAAACTTTT TGTTTTGACT TGTATAATTA 420  
 10 AGGGTTTGA AAGATTCATA ATTCTGAGAG AGGTTTGCAA CCAGGAGATA CAAAGAAGTC 480  
 TCAGTAGTAA TCTGTTCAT GTGCTTTTAC AGCCAGCTAC ATTTAAGGAT GTATTAGTTA 540  
 15 CAGAAATTAT ATGTCTGTGT ATGTGTCTCT ACTCAATAAA GTACATGCCT CCACAAAAAA 600  
 AAAA 604

20

(2) INFORMATION FOR SEQ ID NO: 37:

25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 349 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

GTGAGTGCCC GGGAGCCCCG AGGCCCTGCC CCTAAGAAGG ATATCTYTRA CCGCTCCCTT 60  
 GTCCACACCC TAACCCCCCA GCTGCTCAGG CAGTGGGCAC ATGGCAGGGG CCTCACTGGG 120  
 35 GGCACATAGA GCATTTGGGG GACTGCGAGT GCTCACCTTT GACTTCCTGC AGGTGCGGGG 180  
 AAAACCAGAT CATGATGACC AAAGTYTACA TATTCTTGAT CTTCATGGTG CTGATCCTGC 240  
 40 CCTCCCTGGG TCTCACCAGG TATATGCCAC CACYTTCTGY TCTAAATTCA GAATAAGAGT 300  
 CACATCAGGA GAGCACTGTC CCCAGGANAA TGCAACGGG TTGGCAGCA 349

45

(2) INFORMATION FOR SEQ ID NO: 38:

50 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 672 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

GTAGTCGTTG CGGTTGCCGG GATGGCGAAG ATCTCGCCGT TTGAAGTCGT AAAACGCACC 60  
 TCGGTACCGG TGCTTGTTGG TTTGGTGATT GTWATCGTTG CTACAGAGCT GATGGTGCCA 120  
 60

GGAACGGCAG CAGCGGTCAC AGGCAAGTAA ATAGTAATGC CGGAGCAAGT TTCCTCCGGC 180  
 TTTATCATGT CACCCACTGT GGTATATGCG TTGTGGTCTG CCAACTTTGC CGTGAACAAT 240  
 5 TTCAGCAATA ATCAGATGGC GGCTGGCGCA ATATTCAAGA TAACGCCTGG CAGTGGTGCG 300  
 GCTGATGGTT CAGTGCCTGC GSCACCGTTT YTGCCGTATG TTGCACACCA GGNTCTTTAA 360  
 ACAGTTTTCG SACCGCGTTT AGCGTCAAGG GTTCAATGCC GGTCGGTAGC TCGTCCTTAG 420  
 10 GTTCACCGCG AGCATAAGCA TTAAACATCT CATCAATTTG CTTCTGGCTG GCGCTATCAA 480  
 TACTTTCCAG CATATGTTTA CGCTGGCGGA AACGGGTTAG CGTTTGCCCC ARCMGWTCAT 540  
 15 AGGCAATGGG CTTAATGAGA TAATCAAATA CACCACAACG TACGGCTTCA GACACCGTTT 600  
 CCATATCGCT GGCTGCAGTG GTAAACACCA CGTCGCCGGG ATAATGCCGC TGCACCAGTT 660  
 CATGCAGTAA AT 672  
 20

## (2) INFORMATION FOR SEQ ID NO: 39:

25

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1908 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

30

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

AGAGTTGATA TTTTGTAGAA CAGTAATTTT ACTTTTAAGG AAATTGGCTA GCTCTTTGAC 60  
 35 TNNAGAGCTG TAGGAAGCTC AACATTTCCT TGTAGAGAAC GTTGCTTTTT TTGATTTGTA 120  
 CAGGTATAAA AACATTGCTT TTGTTGAATT GTATAGGTGT AAAAAGGAA TAACTGTATG 180  
 40 CAGGTTTGAA AAGGAAATGT GCTTTAGGCA TGAGTCATAA GATGCCATTG TACTGTAGG 240  
 CATTTTATTT TCCTTTAGAA ATGGACATCA GCTCTTCTCT TCTGACTGGT AACACATAGC 300  
 CCCAAAGCAT GAGATTATTT TTCATTGGGT TTTTATTGTT GTTTAGTTTT GGTTTGTTAC 360  
 45 GCCAGCCGAG TCTGTCTGCG GAACACTGAC TCTGCTCTCT AATGAGAACA AAGTTAGAAA 420  
 TCTGCCGATA ACCTAAAATA ATTTAGAAAT GAATTAAGAA TGTGAAATCG GGTTAAAGTG 480  
 50 ATGATGATAA AATAGCATGC AAGAAACAAG CTCCTTCCAT CAGACTTGGC TACTGTTTTTC 540  
 TTCTGGTACG ATTTGGTTTG GAAGAGCCTC TTGTTTCCTT CTCCTTGGGG TATGTCCTCG 600  
 TTTCTTAATA TGTGTGTAAC ATTATTGAGA TATAATTCAC ATACCTTACA ATTCACTTAT 660  
 55 TTTAAGGGTA CAATTTAGTG GTTTTTAGTG TATTACAAA GTTGTGTAAC CGTGACCACA 720  
 GTCAATTTTA GAACATTTTC TTACCCCAAA AAGAAACCCT GTACCCCTGA GCAGTCACCT 780  
 60 CTCATTTTCT CCCAGTGCCC ACCCCATCCC CGAGCCCKG GAACCACTAA TCTATTTCTC 840

	TCTCTGTAGA TTTGCTTATT CTGGTCATTT CATATAAATG GAATTCTACA ATATTCGGTC	900
5	TTTGGGACT GGCTTCCCAA ATATGATTTT CTATATGGAG TGAGAAAATT CTTCTCATCT	960
	TGAGAACTCT TATTGCTGTG AAAGGGAGTG GTTGGTAAAA TCAATAGATT TCAGGCAAGA	1020
	GGCCAGATA CCTAACAGGT TTTTCTCCGT GAATCTTATG CTGAGTAGTT TTTCTCATA	1080
10	ACCAAGCATT TATGATATAT TACTACTTAT AATACTGTGG CTAGTCTCTA GAATGGATGT	1140
	TGAAATCTTT GCCTCCTCAG TCGGAAGAG TCCTGCTAAA AATCAGGCTA AAAATCAGGC	1200
15	CAAAAATCAG GCCAAATGAC TTGGCAAATA ATTGACAAAG TGGTTTTCAC GTGTGTCTAT	1260
	CTTGTGCTAGC AGCTTGTATA CCTCAGGCCA GGTGAGCTCC CCAAATTTCT TTTTTCATTT	1320
	ACTCCAGTGA GTTCTGCTG TCTTTTTCAA GTATGTACCA TAGGACTTAA AGGTGATTTG	1380
20	GATGCGTTGT AACACTGCTA AATATGCTAA GTACAGAATT TTATCTACAG TACTGTGAGA	1440
	CAGTCAATTA TTGCCTAGGG TAGTTCAAAA ATATGATGTG AGCTAGTTAA GCCTTTGCTT	1500
25	GACTGATTTT AGTGATATTC AGAAGTGTGT ACCAATCAAG GCTCTTTAAA ATACGGAACG	1560
	ACTCACTTAA TAACCAGGA ACCAGCCAAA TACTGTGCAG CCGCAGAATA TGCATATCAA	1620
	TGAGTTGGAG GTGATTATTC TCTGTAATC CTAATGATT GTTTTCTAAG CATTGTGGCT	1680
30	TCTCAGTGGC TTGACAGCAT CTTCTGGTT GTATGTGGCC TGTTTACATG ATGTATTGAA	1740
	TAATGTTGTT TGTGTGAGC ATCAATGCCT GTAACACCAA ACTAAACACG TGTTTTTGGG	1800
35	ATATGTTTCC AATCTTTAAA TGACCTTGCC CTGTCCAATA AATAAATGAT TGTCTACCC	1860
	TGTTAAAAAA AAAAAAATT AAAAAAATG GNGGGGGGC CCGGTACN	1908

40

(2) INFORMATION FOR SEQ ID NO: 40:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 458 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

45

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

	CCTCAAAAAA AAAAANGAAA GGAAAGAGGT CTCTACACAA GCCCGTGATT CTTTATGGCA	60
	AGGATAACA TCAGAAATGT TTCATTTYCK GCTATTAGTT TCCATTCCTT TCCCCATCCA	120
55	GGCATAAAGA GAAACAAAAG ACAATGATGG TATTTCTCTGT GTCTCAGCT TTGGCACTTT	180
	TGTTGATGTT GCTAAGGAGC AGTGACCTTG CTAAAAAGAC TGAATAATCC ACCCACTGAA	240
60	TAGCTAACCT GGGGAGGAAA TGAAAATTTT CTTTGTGGAT CTCCCCAAAT CCATTGTTGT	300

CACCAGGCCC TCCCAGAAC TCCTCAGTTC CTTCACAGTG CAACCCCTGTG TACTTGCCCC 360  
GCAACCCAAT AGTATTGTGC CTCAC TTCAC CTTCATGGG CAACTGCCCT CCCTTCTGGA 420  
5 CATAAACCT CATATTTTAA ATNAAGTTGA AATTTGAA 458

10 (2) INFORMATION FOR SEQ ID NO: 41:

(i) SEQUENCE CHARACTERISTICS:

- 15 (A) LENGTH: 1153 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

20 GGCACAGAGC CTCGACCCA GGTGGTCTGG AGCCTGCCGG GAGAGTGGTG GCATCTGAGA 60  
GGCTGGTCGT GGACTGTGGT TGGGGGAGGT GGGAGCTGTT TTAACCGTGT GCCCCCTCTC 120  
CTGTGCCGGC GTGGGCATCC CCCGGGGCAG TGGAAACCCG GCGCTCCTCC AGCTTCCGAG 180  
25 TCCAGCCAGC CTGGGCGCGG GCGCGCCCC GAGACACCCG AGGAGTCCGT TCCTCCCTGG 240  
TTACGTGGAC TGTGGAGCTG GTCTCTTG TGCTCAGCGC GTGCGGAGGT TGAAGCGTAC 300  
30 CTGCGGAGGT CGCACCAGG CGTGAGGAGG AGGAGGAAG GCATGAGCCG AGCTTGAGGA 360  
ATCCGTGCTC CAACTCTAC ACTCAAGGAT GCACTGCGCA ACTCTGGTGG CGATGGGCTG 420  
GGCAGATGT CCTTGGAGTT CTACCAGAAG AAGAAGTCTC GCTGGCCATT CTCAGACGAG 480  
35 TGCATCCCAT GGAAGTGTG GACGGTCAAG GTGCATGTGG TAGCCCTGGC CACGGAGCAG 540  
GAGCGGCAGA TCTGCCGGA GAAGGTGGT GAGAACTCT GCGAGAAGAT CATCAACATC 600  
40 GTGGAGGTGA TGAATCGCA TGAGTACTTG CCCAAGATGC CCACACAGTC GGAGGTGGAT 660  
AACGTGTTTG ACACAGGCTT GCGGGACGTG CAGCCCTACC TGTACAAGAT CTCCTTCCAG 720  
ATCACTGATG CCTGGGCAC CTCAGTCACC ACCACCATGC GCAGGCTCAT CAAAGACACC 780  
45 CTGCCCTCTG AGCGTCGCTG GATCTCTGGG AGCTCCTTGA TGGCTCCAG ACCTTGGCTT 840  
TTGGGAATTG CACTTTTGGG CCTTTGGGCT CTGGAACCTG CTCTGGGTCA TTGGTGAGAC 900  
50 TTGGAAGGGG CAGCCCCCGC TGGCTTCTTG GTTTTGTGGT TGCCAGCCTC AGGTATCCT 960  
TTTAATCTTT GCTGACGTT CAGTCCTGCC TCTACTGTCT CTCCATAGCC CTGGTGGGGT 1020  
CCCCCTTCTT TCTCCACTGT ACAGAAGAGC CACCACTGGG ATGGGAATA AAGTTGAGAA 1080  
55 CATGAGTTTG GGCTGAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1140  
AAAAAAAAA AAA 1153

## (2) INFORMATION FOR SEQ ID NO: 42:

5

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1983 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

	GGCAGGAGAG GGGCCGAGCC GACAAGATGT TCTTGCTGCC TCTTCCGGCT GCGGGGCGAG	60
15	TAGTCGTCCG ACGTCTGGCC GTGAGACGTT TCGGGAGCCG GAGTCTCTCC ACCGCAGACA	120
	TGACGAAGGG CCTGTGTTTA GGAATCTATT CCAAAGAAAA AGAAGATGAT GTGCCACAGT	180
	TCACAAGTGC AGGAGAGAAT TTTGATAAAT TGTTAGCTGG AAAGCTGAGA GAGACTTTGA	240
20	ACATATCTGG ACCACCTCTG AAGGCAGGGA AGACTCGAAC CTTTTATGGT CTGCATCAGG	300
	ACTTCCCCAG CGTGGTCTA GTTGGCCTCG GCAAAAAGGC AGCTGGAATC GACGAACAGG	360
25	AAAACTGGCA TGAAGGCAAA GAAAACATCA GAGCTGCTGT TGCAGCGGGG TGCAGGCAGA	420
	TTCAAGACCT GGAGCTCTCG TCTGTGGARG TGGATCCCTG TGGAGACGCT CAGGCTGCTG	480
	CGGAGGGAGC GGTGCTTGGT CTCATGAAT ACGATGACCT AAAGCAAAAA AAGAAGATGG	540
30	CTGTGTCCGC AAAGCTCTAT GGAAGTGGGG ATCAGGAGGC CTGGCAGAAA GGAGTCTGT	600
	TTGCTTCTGG GCAGAACTTG GCACGCCAAT TGATGGAGAC GCCAGCCAAT GAGATGACGC	660
35	CAACCAGATT TGCCGAAATT ATTGAGAAGA ATCTCAAAAG TGCTAGTAGT AAAACCGAGG	720
	TCCATATCAG ACCCAAGTCT TGGATTGAGG AACAGGCAAT GGGATCATTC CTCAGTGTGG	780
	CCAAAGGATC TGACGAGCCC CCAGTCTTCT TGGAATTC A CTACAAAGGC AGCCCCAATG	840
40	CAAACGAACC ACCCTGGTG TTTGTTGGGA AAGGAATTAC CTTTGACAGT GGTGGTATCT	900
	CCATCAAGGC TTCTGCAAAT ATGGACCTCA TGAGGGCTGA CATGGGAGGA GCTGCAACTA	960
45	TATGCTCAGC CATCGTGTCT GCTGCAAAGC TTAATTGGCC CATTAATATT ATAGGTCTGG	1020
	CCCCCTTTTG TGAAAAATATG CCCAGCGGCA AGGCCAACAA GCCGGGGGAT GTTGTTAGAG	1080
	CCAAAAACGG GAAGACCATC CAGGTTGATA AACTGATGC TGAGGGGAGG CTCATACTGG	1140
50	CTGATGCGCT CTGTTACGCA CACACGTTTA ACCCGAAGNT CATCCTCAAT GCCGCCACCT	1200
	TAACAGGTGC CATGGATGTA GCTTTGGGAT CAGGTGCCAC TGGGGTCTTT ACCAATTCAT	1260
55	CCTGGCTCTG GAACAACTC TTCGAGGCCA GCATTGAAAC AGGGGACCGT GTCTGGAGGA	1320
	TGCCTCTCTT CGAACATTAT ACAAGACAGG TTGTAGATTG CCAGCTTGCT GATGTTAACA	1380
60	ACATTGGAAA ATACAGATCT GCAGGAGCAT GTACAGCTGC AGCATTCCTG AAAGAATTCTG	1440

TAACTCATCC TAAGTGGGCA CATTTAGACA TAGCAGGCGT GATGACCAAC AAAGATGAAG 1500  
 TTCCCTATCT ACGGAAAGGC ATGACTGGGA GGGCCACAAG GACTCTCATT GAGTTCTTAC 1560  
 5 TTCGTTTCAG TCAAGACAAT GCTTAGITCA GATACTCAAA AATGTCTTCA CTCTGTCTTA 1620  
 AATTGGACAG TTGAACTTAA AAGGTTTTTG AATAAATGGA TGAAAATCTT TTAACGGAGA 1680  
 CAAAGGATGG TATTTAAAAA TGTAGAACAC AATGAAATTT GTATGCCTTG ATTTTTTTTT 1740  
 10 CATTTACAC AAAGATTTAT AAAGGTAAAG TTAATATCTT ACTTGATAAG GATTTTTTAAG 1800  
 ATACTCTATA AATGATTAAA ATTTTITAGAA CTTCTAATC ACTTTTCAGA GTATATGTTT 1860  
 15 TTCATTGAGA AGCAAAATG TAACTCAGAT TTGTGATGCT AGGAACATGA GCAAACTGAA 1920  
 AATTACTATG CACTTGTGAG AAACAATAAA TGCAACTTGT TGTGCAAAAA AAAAAAAAAA 1980  
 AAA 1983  
 20

(2) INFORMATION FOR SEQ ID NO: 43:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1406 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

ATGATGATGA CTTTGAAGAC GATTTTATTC CTCTTCCTCC AGCTAAGCGC CTTGAGGTTA 60  
 35 ATAGTTGGAA AAGACTCTAT AGATATTGAC ATTCTTCAA GGAGAAGAGA AGATCAGTCT 120  
 TTAAGGCTTA ATGCCTAAGC NCTTGGTCTT AACTTGACCT GGGATAACTA CTTTAAAGAA 180  
 40 ATAAAAAATT CCAGTCAATT ATTCCTCAAC TGAAAGTTTA GTGGCAGCAC TTCTATTGTC 240  
 CCTTCACTTA TCAGCATACT ATTGTAGAAA GTGTACAGCA TACTGACTCA ATTCTTAAGT 300  
 CTGATTGTG CAAATTTTTA TCGTACTTTT TAAATAGCCT TCTTACGTGC AATTCTGAGT 360  
 45 TAGAGGTAAA GCCCTGTGT AAAATAAAGG CTCAAGCAAA ATTGTACAGT GATAGCAACT 420  
 TTCCACACAG GACGTTGAAA ACAGTAATGT GGCTACACAG TTTTTTTAAC TGTAAGAGCA 480  
 50 TCAGCTGGCT CTTTAATATA TGAATAAACA ATAATTTAAA ACAAATCATA GTAGCAGCAT 540  
 ATTAAGGTT TCTAGTATGC TAATATCACC AGCAATGATC TTTGGCTTTT TGATTTATTT 600  
 GCTAGATGTT TCCCCCTTGG AGTTTGTGCA GTTTCACACT GTTGTCTGGC CCAGGTGTAC 660  
 55 TGTTTGTGGC CTTTGTTAAT ATCGCAAACC ATTGGTTGGG AGTCAGATTG GTTCTTAAA 720  
 AAAAAAAAAA AAAACGACAT ACGTGACAGC TCACTTTTCA GTTCATTATA TGTACCGAGG 780  
 60 GTAGCAGTGT GTGGGATGAG GTTCGATACA GNCGTATTTA TTGCTTGTCA TGTAAATTAA 840



AAACCTTGTA TTAACTCTT TTCAATCCTT TTAGATAAAA TTGTCTTTG CAAGAATGAT 900  
 TGGTGCTTAT TTTTCAAAA ATTTGCTGTG AACACGTGA TGACAACAAG CAACATTTAT 960  
 5 CTAATGAAC ACAGCTATCT TAATTTGGTT CTCAAGTTT TCTGKTGCAC TTGTAAAATG 1020  
 CTACAAGGAA TATTAAAAA ATCTATTCAC TTAACTTAT AATAGTTTAT GAAATAAAAA 1080  
 10 CATGAGTCAC AGCTTTTGT CTGTGGTAAC CTATAAAAAA AGTTTGTCTT TGAGATTCAA 1140  
 TGTAAGAAG TGAAACAAT GTATATGTTG TAAATATTTG TGTGTTGTA GAAATTTTGT 1200  
 TCATAAGAAA TTAAGAAG TTACCAGGAA GGTTTTAAAG TTAGAAATAT TCCATGCCAA 1260  
 15 TAAATAGGA AATTATAAAT ATATAGTTT AAGCCTGCAT CAGTGGGAGT CTTGGCTATG 1320  
 TAGTTATGTA GTTATTATGN AACCACCAAG ATTTTMTTGG CTATTTACCG TAACCAAAGG 1380  
 20 GGCCGATTAA NTGGTTTGAA GNCTTG 1406

25 (2) INFORMATION FOR SEQ ID NO: 44:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1391 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

35 GGGCTGAAG GCGCRGCC AGTCCGAGC AGTGCTCGCT CCTGCTGGG GCGCTGGGC 60  
 CCCGGGCGTC GCCATGACCA GTGAGCTGGA CATCTTCGTG GGAACACGA CCCTTATCGA 120  
 CGAGGACGTG TATCGCTCTT GGCTCGATGG TTAATCGGTG ACCGACGGG TGGCCCTGCG 180  
 40 GGTGCGCTCG GGAATCCTGG AGCAGACTGG CGCCACGGCA GCGGTGCTGC AGAGCGACAC 240  
 CATGGACCAT TACCGCACCT TCCACATGCT CGAGCGGCTG CTGCATGCGC CGCCCAAGCT 300  
 45 ACTGCACCAG CTCATCTTCC AGATTCCGCC CTCCCGGCAG GCACTACTCA TCGAGAGGTA 360  
 CTATGCCTTT GATGAGGCCT TTGTTCGGGA GGTGCTGGG AAGAAGCTGT CCAAGGCAC 420  
 CAAGAAAGAC CTGATGACA TCAGCACCA AACAGGCATC ACCCTCAAGA GCTGCCGGAG 480  
 50 ACAGTTTGAC AACTTTAAAC GGGTCTTCAA GGTGGTAGAG GAAATGCGG GCTCCCTGGT 540  
 GGACAATATT CAGCAACACT TCCTCCTCTC TGACCGGTTG GCCAGGACT ATGCAGCCAT 600  
 55 CGTCTTCTTT GCTAACACC GCTTTGAGAC AGGGAAGAAA AACTGCAGT ATCTGAGCTT 660  
 CGGTGACTTT GCCTTCTGCG CTGAGCTCAT GATCCAAAAC TGGACCCTTG GACCCGTGCA 720  
 60 CTCACAGATG GATGACATGG ACATGGACTT AGACAGGAAT TTCTCCAGGA CTTGAAGGAG 780

CTCAAGGTGC TAGTGGCTGA CAAGGACCTT CTGGACCTGC ACAAGAGCCT GGTGTGCACT 840  
GCTCTCCGGG AAAGCTGGGC GTCTTCTCTG AGATGGAAGC CAACTTCAAG AACCTGTCCC 900  
5 GGGGGCTGGT GAACGTGCCG CCAAGCTGAC CCACAATAAA GATGTCAGAG ACCTGTTTGT 960  
GGACCTCGTG GAGAAGTTTG TGAACCCCTG CCGCTCCGAC CACTGGCCAC TCAGCGACGT 1020  
GCGGTTCTTC CTGAATCAGT ATTCAGCGTC TGTCCAATCC CTCGATGGCT TCCGACACCA 1080  
10 GGCCCTCTGG GACCGCTACA TGGGCACCCT CCGCGGCTGC CTCTGCGCC TGTATCATGA 1140  
CTGAGGTGCC TCCCAACGTC CGCCACGCT GACAATAAG TTGCTCTGAG TTTGGAGACT 1200  
15 GGTCTCTGCT CCGGGGAGCA AGTGGGGGGC GTGCAGATGT GCCTGTGTCT GTCTCTGAGC 1260  
ACCTGGTGTC CGTGTAACG GATGGATGTG TNCNGTGGCT CCTTGGGAAC TGAGACATAT 1320  
20 CTCAGGGAAT GGTGTCTGTG CTCAGCCCAT CCACCAGAAG AGTCTGCTCA CAAAAAAAAA 1380  
AAAAAAAAA A 1391

25

(2) INFORMATION FOR SEQ ID NO: 45:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1569 base pairs  
30 (B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

GGCAGAGTG GAGATGGCTG CGGCCGTGGC GGGGATGCTG CGAGGGGGTC TCCTGCCCCA 60  
GGCGGGCCGG CTGCCTACCC TCCAGACTGT CCGCTATGGC TCCAAGGCTG TTACCCGCCA 120  
40 CCGTCGTGTG ATGCACTTTC AGCGGCAGAA GCTGATGGCT GTGACTGAAT ATATCCCCC 180  
GAAACCAGCC ATCCACCCAT CATGCCTGCC ATCTCCTCCC AGCCCCCAC AGGAGGAGAT 240  
AGGCCTCATC AGGCTTCTCC GCCGGGAGAT AGCAGCAGTT TTCCAGGACA ACCGAATGAT 300  
45 AGCCGTCTGC CAGAATGTGG CTCTGAGTGC AGAGGACAAG CTTCATTATTG CGACACCAGC 360  
TGCGGAAACA CAAGATCCTG ATGAAGGTCT TCCCCAACCA GTTCCTGAAA GCCCTTCCTG 420  
50 GAGGATTCCA AGTACCAAAA TCTGCTGCCC CTTTTGTGG GGCACAACAT GCTGCTGGTC 480  
AGTGAAGAGC CCAAGGTCAA GGAGATGGTA CGGATCTTAA GGGACTGTGC CATTCCTGCC 540  
GCTGCTAGGT GGCTGCATTG ATGACACCAT CCTCAGCAGG CAGGGCTTTA TCAACTACTC 600  
55 CAAGCTCCCC AGCCTGCCCC TGGTGCAGGG GGAGCTTGTA GGAGGCTCA CCTGCCTCAC 660  
AGCCCAGACC CACTCCCTGC TCCAGCACCA GCCCCCAG CTGACCACCC TGTTGGACCA 720  
60 GTACATCAGA GAGCAACGCG AGRAAGGATT CTGTCATGTC GGCCAATGGG AAGCCAGATC 780

	CTGACACTGT TCCGACTCG TAGCCAGCCT GTTTAGCCAG CCCTGCGCAT AAATACACTC	840
5	TGCGTTATTG GCTGTGCTCT CCTCAATGGG ACATGTGGAA GAACTTGGGG TCGGGGAGTG	900
	TGTTTGTAC TTGGTTTCA CTAGTAATGA TATTGTACAG TATAGGGCCA CTTGGAGATG	960
	CAGAGGATTC CATTTTCAGAT GTCAGTCACC GGCTTCGTCC TTAGTTTTTCC CAACTTGGGA	1020
10	CGTGATAGGA GCAAAGTCTC TCCATTCTCC AGGTCCAAGG CAGAGATCCT GAAAAGATAG	1080
	GGCTATTGTC CCTGCTCTCC TTGGTCACTG CCTCTGCTG CACGGGCTCC TGAGCCCACC	1140
15	CCCTTGGGGC ACAACCTGCC ACTGCCACAG TAGCTCAACC AAGCAGTTGT GCTGAGAATG	1200
	GCACCTGGTG AGAGCCTGCT GTGTGCCAGG CTTGTGCTG AGTGCTGTTA CATGTATTAG	1260
	TTCTTTTACT GCTGACCACA TTGTACCCAT TTCACAGAGA AGGAGCAGAG AAATTAAGTG	1320
20	GCTTGCTCAA GGTTCATGCAG TTAGTAAGTG GCAGAACAGG GACTTGAACC AAGCCCTCTG	1380
	CTCTGAAGAC CGCGTCCTGA ATTTCTTCAC TAGAGCTTCC TCATCAGGTT ACCCAGAAGT	1440
25	GGGTCCCATC CACCATCCAG GTGTGCTTGG ATGTTAGTTC TCCACCCTCG AGGTGTACGC	1500
	TGTGAAAAGT TTGGGAGCAC TGCTTTATAA TAAATGAAA TATATTCTAA AAAAAAAAAA	1560
	AAAAAAAAA	1569

30

## (2) INFORMATION FOR SEQ ID NO: 46:

35

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1924 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

40

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

	GGGCCCCCCC WCGWKTMTT TTTTMTT TTTAATTAGG ATAATGCCTT TATTAACGAG	60
45	AATGAAACGT TCATTCTCTC TTCCACTCCT TCTCGTTGGT TTTCTGGACA CAGCTCACCT	120
	GATCCTGCTA GAAACGTTGT CAGTCGCTT GTGGCTTCCC TCCTTGATTG ACTCACGCTG	180
50	TGTGATGTCT TGAGAAGTAT CTATCCACTT CATGTGAATG AGCACTCCAA TATCAGCCAA	240
	CATCAATCAT TCTTACCTAA AGAATAATAA GAAAAAGTTA ATATAAAGA CAAGGGTATA	300
	AAATAAAGGT TTGAAAATGC TAGTCAACTT CAAAATTAA AGAGTAAAAA TCCAGAGATA	360
55	AAGATTGGGG GTAAGTTACA GCATAAAAAA ATAGGAAGAA ACTTCATGGT GGGGGGAAA	420
	TCTAAAATTA TTCTTACATA AAATAAGTAG ACACCTGAAT TAGAATGAAA ACTGTATTTT	480
60	CTTTAAATG TAAAGCCTG ACTCTCAGTT TCACCAGTCT GAGCACAAGT TTGACTGCAA	540

	CCCAAAATAT ACTATCCCTT ATGTGAAGGT ATGTGACAAC GTTGACCTCA CCAAATGAGT	600
	TTTAACATCA GCTCTTTTTT CATATGAAAG CACATACCCT GCTCCCCATT CAAGTATGTC	660
5	TTCCATGTC AGGCAGGCTG ACCACCTTCA GCAGGAGTCC TCCAAGAGTG CCCAACTCCC	720
	CTTCCCACAG TACACAACGC TGTAGTTGTT GTCCTGCAAT CCTTTGTATT TACCTCATTC	780
10	TTTCCCATCT AAGTCCTCAC TGAGTTTAA AGTTAGGGCT GGAAAAGCTA TGCCTTACTG	840
	GGACAGCAAG GAACCAATTT TTTTCTGAGG GAGAAGACAT TCACCTTAC TATATGCCTG	900
	GCAGGGCCAC AGTGACAAAA ACAAGATCA GCCTTCATTC AAGTTCCAGG TTTTCTTCC	960
15	TCCCTGAATG ATTACTGCAA AGGGTATATG AAGTAAGAGT TCCCTGTTGC ACATGTACCA	1020
	TCCATAAGGG ATACTATATC GTTTTGCAAT CTCCCCCA TTCTCCACAT TGTCTATCT	1080
20	TAAGTCCAAG CCTTTTTCAC TCTCAAAAAA AAAAAAAAAA TATTTTTTTC AGCACTGGTG	1140
	TTCAAAGCA ACGTTTTTAT GGTTAATGGT TTACCAGCAA CTGTTGAGAT TTCCAGTTGA	1200
	GTCTTAAAAA TTGCCAATCA TTATCTAGCA GCAATGACAG ATGATTAGGA GCAGTCAAAT	1260
25	CCTCTGAATT CTTCCCTAA TAGGCAGCCA TTTGAGAACT GCACTAGCTG ACATCACTAA	1320
	AACATTATCA GCTAAAGCCA AAACCAAATA AAGGCCCAGA CCAACATCCT GGCTCTCTAA	1380
30	AACCTGTCCA AAATCATTA GTGAAAGCA GTAAATGCAG GACTGTGGAT CATGTCACTG	1440
	CAGCTGACAA TGATTAACAA TAGGAGACAT GCAACCCCA TTAAGGTTAA AAGTCCAAAA	1500
	CTAGTCACAC GCATCTCTT ATTGGGGAAA AGTGAGACTA TTATGCATTG TTGGTAGGTT	1560
35	TGCAACCTTG CATGAAGAGC ACCCATTGCA TTTCTTTCAT CTTTCAGAAA GCACCGGTAT	1620
	CTGTTCCAAG GGCCTAACAG TACGAAAATA CATTCTGGCA TCACACCTCT GAACCCAAGA	1680
40	CTGTTCTCAT TAAAAATAAT TTTGGTTTGT AACAAAATTA TGAAATACAA TGCAAGCACC	1740
	TCGGTATAGC ATTATTAAGT AAACCACTTA ATTCCCAGCT TTTTGAGTTT TTTAAAAAAA	1800
	CCCACTGCAC TAAGATTCAC AATTCATTGC TACATACAAA TTAAAGCTAG TAAGAACACA	1860
45	CTAACGTCAC AAGTTTCTCA TTCTAAAGTG CAAAAGCCTA ATCATCTGAA AGTGAACAGG	1920
	GTAA	1924

50

(2) INFORMATION FOR SEQ ID NO: 47:

55

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 475 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

TGGTGTGGGG CCCAGAAAMC AAGGGACCAG TGAAAACAMC CCCAGAGACT TGTATCCGCC 60  
 AGGAAAGCCA TTGCCAMTYC TGAGCCCTTG AAGGGCAAGG AGGGAAACAG TGTACCAGA 120  
 5 GCCAGTAAG AACTGCTGTC ATGAAGGAGG GGCCACCTTG TAAGAGACAT CATTACTACC 180  
 AGAACTGTGG TGCCAAATG CTGGTGCTC TCTTTGGAGA AACCAACCAG ATACATCTGC 240  
 10 TGGAGACCCA GGTGGGCACA GAGAAGGGTG GAGAGAGAAT CTGGGAAGAG AAATGGAGAA 300  
 TAAGCAGCAC AGTGTATTC ATTTCTGTAA ATTCTATGT AGAAGGCTCA GTGTTAGAAA 360  
 TAAAGTTATT CTACTAGTTG CAAGTTAAGT GTTCTGTGTT GTTCTGCTTT CCTGTTAGCA 420  
 15 TAAGTAACT CCCTTTGGAA CTACACAGGT ATGTCTCTCC TTCAACATGT GTGAA 475

20 (2) INFORMATION FOR SEQ ID NO: 48:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 346 base pairs  
 25 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

AAGGGACAGA GACCTGGATT CAGATCTCAT TTTACAATGA AGACCCCAAT GCAGAAAGTC 60  
 ATGTCTGAAA TTCTGAGCTT ACTCTTCTGC CTGCTGGGAC CTGCTCTGGA TGAGAGAAGG 120  
 35 GAGGAAAAGG ACTAATCAGA GGAGCCAATG AAGTCACTCC ATGAGTTTCC TGAACCCCTGC 180  
 CCAGCTAGAG ATTAACGTYT GACCWTC AAC GTAGGACACT GTGCAGATGG CTACTTGCTG 240  
 GCGCACATGA AGACCAAAGC CAGGACCAAG CCCCMASCCT GCTWAACACG GCAGARTCTT 300  
 40 GCCCAGCCMA CYTCTGTGAR AATCTGCTTC CCTCCACAGC TGACCC 346

45 (2) INFORMATION FOR SEQ ID NO: 49:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1366 base pairs  
 50 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

TAGGTGTCAG CCGCCACCCC CCCCCATAT GCAGATTTAC TSGGCATGGT AGTGGCCAGC 60  
 TTCTAACACA GCTGGTATTT CAAGTCTCCT GGGACCTCAC TCAGGAATGA TACCCCCTCA 120  
 60 GTAGAAGCAG CAGGTGATCT TAACTCCTTT CAAAGAGCAG GCCTGTCTGG GAAGCCATGT 180

5 CCTCAGCAGG CACAGCAACC CCTCTGGAAA TGGATCACAA ACTCACTTCT CAGCCAGGCA 240  
 GGCCAAGCTT CTATGTGAAC AGTAGGCACA GTATAGTCGG ATCATCACAT CAGCTGGGTT 300  
 TTGGTMTAG TCATCTAGAG TCGTCTGGAC TAAAGGTCTT TCAGGTCTCC TTGCCCTGTG 360  
 AGTGCGTGAA CCTCCCCACC CGAATTGCCT CAGTTGTCTT GAGCCTCATG TCTCTCCTGG 420  
 10 TGGTGGGCCA GGCCCTGCA TGGGAAGGGA GCCTGCTGCG GGGCAGGCCA GCTGGGGGTG 480  
 CTCACCTATG CGCAATGANA GTTATTGAAG GACTGGTTGT TGATGTTGGT GAGCGTATCC 540  
 TTCATGGCCA GCGCGAAGTC GGCCAGGTCA GCCAGGTGCT GCCAGCGCTC TCTCTCGGAC 600  
 15 TTGTCTTCCT GTGCCAGGGG ACCGTGGAGA AAGTGTCAAG GGGCGCTCAC TGCAGCAGCC 660  
 TGCTCTGCTG CCTTCCCTGG CAGTGTCTG GGGGTGGATT CCCTACAMCT AGATGTTCAA 720  
 20 GGCCTTACTT TTCTCTCCAC AAAGGAGTCG CAGCCACGCT AGCTCTGACT TGCCACTGTG 780  
 ACAAGTTCA CGTAGCAGGT CTAGGCAAAG ACTGGGCAAT TGAGCAGAGG AGACGGACCT 840  
 25 GTGAGTCTGA CCRYGAGSCG GRCCCTTCA CCTTGGCTGG GCTGGTCTG GTCCTTAGGT 900  
 TTTGTCAAGT TGTCCTTGTT TGGATCCCTC AACTAGGTGA TAAGCACTGG AGGGGGATGA 960  
 CCCGCTTGG ACGTGTCTTCT TTAACCTCAT CCATATAATA GGGCGTGGG ATGGTTGTAG 1020  
 30 AGGTAAAGCA GGATGATGGT GTTTTAAGAC CAGAGCTTGG GACCAGGGCT CCTACACCTA 1080  
 ATTTCTCTC CTGGTAGCTG AACAAAGGTC TAAATTAGCT TAACAAAAGA ACAGGCTGCC 1140  
 35 GTCAGCCAGA GTTCTGAAGG CCATGCTTTC AGTTTCCCTT GTTGACAATT GCTCTCCAGT 1200  
 TCCTATGAAA GCACAGAGCC TTAGGGGGCC TGGCCACAGA ACACAACCAT CTTAGGCCTG 1260  
 AGCTGTGAAC AGCAGGGGGT TGTGTGCTG TTCTGTTTCT CTGTTTGGG AACTTTCTCA 1320  
 40 ATAAACCTA TTTCTTATTT ATAAAAAAA AAAAAAAA AAAAAA 1366

45 (2) INFORMATION FOR SEQ ID NO: 50:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1405 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

55 GCAGTAATTC CTGTTAGCCA CTGCATCCAC CAAACTAGT TTATTTTTC CCTCAAATTC 60  
 ATGATTTTFA CGTCTGTAC AAAGGAATT TTGCTGATAG CTCTTTGGGT CCCACTGTTC 120  
 60 CATTITATGC TAATAGATTC CATCTAGGG CCCAGCGTC TCTTGACTGA TGGTGTTCCT 180

	TTTAACCCCTT GGCATGTATA ATAGAATTTT GGTGAATGAA AGAACCCAAA TAGGCCAGAT	240
	AGTCCCCCCA GGCCTTGATA TCCATAAAAG GCTTGGGAAT GCATTATGTA ATTGTCCTTA	300
5	GTCTTTTGT TGTTTTAGAA AAAAAAACA AGATGGGCTC AGATGGATGC CTACGTAAAA	360
	ATGGTTCCTA GCTGTGTACT CATAACTTTT CTTTGAATTG AGTAGTGAAA GGAAGGAGGA	420
	GGAAAGGAAA TTAAATGTCC TTCTAGTATT CTCTGGACTC AAGTCTGACA TATGAGATAA	480
10	TAACCTATAT TGAAATGCCA AGAATGTAT CTGAAACAAG AGAACAGTTT GACACATTTA	540
	TCATGCCTTC ATATTACATA TTAAC TGAAA CCAATTAATA AACATATGAA ATATCCATTG	600
15	CACAAGGCAA AGGCACCTAA ACCTTTTGT TCTTTTCTA CATAGCAGAA ATTGATTTTT	660
	TTTTTATTTT TTTAGGGGAA CCTATATAAT TATGACCCAG TGATGTCTTT TGGTGACTTA	720
	AGCTTATGAA TTCAGGTAC AATTGAGTTG ATTCTAGATG GTTACTACCT TGAAAAGGAT	780
20	GTGGTGCCT TATGTGACAC GAGCCAGAGC CTGCTGGGA ATAAACAAAG CAGGTTTCAT	840
	GCCAACACCA ACTCGTAGCT TTAGTGGGCA GATGGGGAGT GGTTCACAGA CTTCCCAAAA	900
25	TGTGGGGCT TTGGGATTTT CCACACCATC CCACGTGTGT TGTTCATTCT TCCTCTTTTC	960
	ACACTCTTGG ATGGATWATT TGRAAATGGT GRAAWYMMCY YYKRAATTG CCCAATAGCC	1020
	WTGRCCACC ATTCTTWATG ACACCATAAC CAAATAGTTC CWTAAATGTTG AAATATTAGA	1080
30	AACCTGTTAC CAGCCYKSM KTWACCCWVA WPTTCCCAT GTTGTGGAA TTGATATTGA	1140
	AATAGCAGG CTAAGGAATT ACTGGCAAGT TTTAGCCTGT GGTAAATACC TTAGGTTAT	1200
35	TTAAATATT GTAATTTTAT TTAAATGTTT ATGAATGTTT GAAAGGAACA AAATTATCAG	1260
	GGATGGCTCT TTGCCATGGG TCTTATTTTC ACCCTCTTTT CTGTAAGAAA AAAGAACAAT	1320
40	GTCTTAATGT ATTTTAAAG TTTTGGTAT AGTTTCTAAT TCCAATTTTA ATAAAAGTTT	1380
	TWTRTAAAAA AAAAAAAAAA AAAAA	1405

45

(2) INFORMATION FOR SEQ ID NO: 51:

(i) SEQUENCE CHARACTERISTICS:

50

- (A) LENGTH: 504 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

55

CGGATTTTCT AGGACCCCAA AAAAAAAAAA AGGNAAAAA AAACCCNCAA AACCANCCAA 60

AACCCAAAA AAAAAAAAAA TCCACAAAA CAAAAAACT ATAAAAAGA AAGAATTAAA 120

60

AACTTTCAGA GAATTACTAT TTACTTTATT AACTTACGGA TTTATTATAT AAATATATAT 180

5 TCACCTAGCA ACATATCTCT GCCGTCTCTC CTGCTCTCAT AATGAAGACA TAGCCGATTC 240  
TCTGCCCCGG CCCCTTGCTG ATGCTCTCTC GGGTCTGCGT CGGGCGTGGG TCTCTGGGGA 300  
CCCTCCAGAG GTGGAGGTGG GCTGATGGCC TGGCTGCTG GTGGTTGATG GTTTTGCTCC 360  
CCCTACCTTT TTTTITGAG TTTATCTGA TTGATTTTTT TTCTTGGTTT CTGGATAAAC 420  
10 CACCTCTGG GGACAGGATA ATAAACATG TAATATTTTT AAGAAGGAAA AAAAAAAAAA 480  
AAAAAACTNG GGGGGGCCCC CGAA 504

15

(2) INFORMATION FOR SEQ ID NO: 52:

20 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 777 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

NAAGTATCTT GCCAGTTTA TTACAGAGGA CGATAAATGA TTCCATGTGG ATAGGGCATA 60  
ACATACAGAG AATGAGACTA TGCCAGAAAT GGGAGGAGGC ATTTGAAACA ACATGAGTAT 120  
30 CTCAGGGACA GATGGATTGA TTCTGCTATT GGTAGGCTG GAAGCAANGG TCAGAAGTAG 180  
CAAAAAATGG ATACCAAAG CACTATTWGT CACCAAGCT AAGTGAATA GCTGGCCCAG 240  
35 TAGGAGAAAT GCAGGTTTTG CTCTACACTA AGTTCTCAA CTCTTGATAA GCCTCCAAAA 300  
ACAAATGTTA GGGGAAAAA ACGCAGCTGG TTATGAAAAG ATATATCTCA TTTCATTAAA 360  
AAATCAATGT CAATGCTGTT AATAGAATCC TTTTATCTTC AGGACAGAGG CAATGCCCTA 420  
40 AACAAACACC AGCTCAAGAG CCTCTGATGC CAACCTAGAG GGTACCCAAA CACAACTTA 480  
GCATAGAGGT AAGAATCTCT ATGCTTTTTG GTGGAGGCAA AGCCATTTGG TTGGTACTTC 540  
45 ACAGGAACAT CTTCTACCA AGTCTTCATC ATATGGTATG TGCCACGAGT CTCCAGTTGT 600  
TTGCACCACT GTGTCATAGC TGAGAATACG CTGAAAGGTT AGTTTIGATC CTGGAACCT 660  
ATTTACAATT GCCAGCTGAT GTCCCTGCTG CCACTTAAAA AAGGCTTGGG TCTGGCATAG 720  
50 GCAGAMAGGC CTGTGGTCCC CTCGTGCCGA TTCTNGGCTC GAGGCCAATT NCCTTAT 777

55

(2) INFORMATION FOR SEQ ID NO: 53:

60 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 602 base pairs  
(B) TYPE: nucleic acid



(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

5  
ATGACTACAG TGTATACCC TCCAATCTTT GCAGGTGGGC ATGGAACACT GCTGTATCA 60  
CTCTGTGCAC GGTATAAATC CATATATCCA CAAAAACACA CATCCATCCA TCAACATATA 120  
10 CATGGTTTGG GATGAGCAGG TCAATAGTTT TGAGAGGGAG TTTGTTCTT TTTTTTTTCT 180  
CATTATACTC TTAAATTGTT GTCAGTTATC AAACAAACAA ACAGAAAAAT TGTTTGGAAA 240  
AACCTTGCAT ACGCCTTTTC TATCAAGTGC TTTAAAATAT AGACTAAATA CACACATCCT 300  
15 GCCAGTTTTT TCATTACAGT ACAGTATCCT TACCTGCCAT TTAATATTAG CCTCGTATTT 360  
TTCTCACGTA TATTTACCTG TGAATTGTAT TTGTTATTTA AACAGGAAAA AAAACATTCA 420  
20 AAAAAAGAAA AATTAAGTGT AGCGCTTCAT TATACTATTA TATTATTATT ATTATTGTGA 480  
CATTTTGGAA TACTGTGGAA GTTTTATCTC TTGCATATAC TTTATACGGA AGTATTACGC 540  
CTTAAAAATA CGAAAATAAA TTTTACAAGG TTCCGGTTTT GGTGGTGGAA AGAGTAAATT 600  
25 GA 602

30

(2) INFORMATION FOR SEQ ID NO: 54:

(i) SEQUENCE CHARACTERISTICS:

35 (A) LENGTH: 1749 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

40 AGTCACTGAC TTGGAGCCGC TCGGGGAAG TCCCGCCAG ACAGGCGGTG GGTGGGAATG 60  
CCTCACTTCA GTTTGAAGAG GGTCCGATC CAAAGGGTT AAAACGAGCG AACCCCGATC 120  
45 CCCGACCACA CTTCGCCCT CCCTAAAACG CACACCCGC TAGCCATGGG CAGCCGCGAC 180  
CACCTGTTC AAGTGCTGGT GGTGGGGAC GCCGAGTGG GCAAGACGTC GCTGGTGCAG 240  
GATTATCCC AGGACAGCTT CAGCAAACAC TACAAGTCCA CGGTGGGAGT GGATTTTGCT 300  
50 CTGAAGGTTT TCCAGTGGTC TGACTACGAG ATAGTGCGGC TTCAGCTGTG GGATATTGCA 360  
GGGAGGAGC GCTTCACCTC TATGACACGA TTGTATTATC GGGATGCCTC TGCCTGTGTT 420  
55 ATTATGTTT ACGTTACCAA TGCCACTACC TTCAGCAACA GCCAGAGGTG GAAACAGGAC 480  
CTAGACAGCA AGCTCACACT ACCCAATGGA GAGCCGGTGC CCTGCCTGCT CTTGGCCAAC 540  
AAGTGTGATC TGTCCCTTG GGCAGTGAGC CGGGACCAGA TTGACCGGTT CAGTAAAGAG 600  
60

	AACGGTTTCA CAGGTTGGAC AGAAACATCA GTCAAGGAGA AAAAAATAT TAATGAGGCT	660
	ATGAGAGTCC TCATTGAAAA GATGATGAGA AATTCCACAG AAGATATCAT GTCTTTGTCC	720
5	ACCCAAGGGG ACTACATCAA TCTACAAACC AAGTCCCTCCA GCTGGTCCCTG CTGCTAGTAG	780
	TGTTTGGCTT ATTTTCCATC CCAGTTCCTG GAGGTCTTTT AAGTCTCTTC CCTTTGGTTG	840
10	CCCACCTGAC CATTTTATTA AGTACATTTG AATTGTCTCC TGACTACTGT CCAGTAAGGA	900
	GGGCCCATTG TCACTTAGAA AAGACACCTG GAACCCATGT GCATTTCTGC ATCTCCTGGA	960
	TTAGCCTTTC ACATGTTGCT GRCTCACATT AGTGCCAGTT AGTGCCTTCG GTGTAAGATC	1020
15	TTCTCATCAG CCTCAATTT GTGATCCGGA ATTTTGTGAG AAGGATTAGA AATCAGCACC	1080
	TGCGTTTTAG AGATCATAAT TCTCACCTAC TTCTGAGCTT ATTTTCCAT TTGATATTCA	1140
20	TTGATATCAT GACTTCCAAT TGAGAGGAAA ATGAGATCAA ATGTCATTTT CCAAATTTCT	1200
	TGTAGGCCGT TGTTCAGAT TCTTCTGTG TTGGAATGTA AACATCTGAT TCTGGAATGC	1260
	AGAAGGAGGG GTCTGGGCAT CTGTGGATTT TTGGCTACTA GAAGTGTCCC AGAAGTCACT	1320
25	GTATTTTTGA AACTTCTAAC GTCATAATTA AGTTTCTCTT GTCTTGGCAT CAAGAATAGT	1380
	CAAGTTTTTT GGCCGGGCAT GGTGGCTCAT GCKGTAATC CCAGCACTTG GGGAGGCCAA	1440
30	GGCAGGCGGA TCACATGAGG CCAGGAATTC GAGACCAACC TGGTCAGCAT GGCAAAACCC	1500
	CGTCTCTACT AAAAGTACAA AAATTAGCCA GCGTGATGG CACGTGTCTG TAATCCCAGC	1560
	TACTCTGGAG ACTGAGGTGG GAGAATCGCT TGAGACTGGG AGGCAGAGGT TGCAGTGAAC	1620
35	CGAGATCATG CCACCGCACT TCAGCCTGGG TGACAGAGAA GGACTCCGTC TCAAAAAAAA	1680
	AAAAAATAA AAAACTCGAG GGGGGGCCCC GTACCCAAAT CGCCSTGATA GTGATCGTAW	1740
40	ACAATCNAA	1749

(2) INFORMATION FOR SEQ ID NO: 55:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1896 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

55	AAAGAGATGG GCTCTTTATT TTCTCGAAAA ACCAATTTGG AGTTACTCAT TTTTCCATAA	60
	CATFAAATTT CTTACAGTGA ACTACATATT GTCCATAAGT GCTTCATCAG GACTCATCGC	120
	CCTCCTGTCT ACTGGCTCCA AATAGACCAT GTCAGCTTCA CCCCCTGGCT TTGTGTCTAT	180
60	GGGTGGCCTG TGGTATATGG AAAAGTAGCA GGGTGGTCAG GGTGGGAGAC ACAAGATGTT	240

	TTTATAGTCT AGAGCCTTTA AAAAACCAG CAGAATGTAA TTCAGTATTT GTTTATGGC	300
5	TGTTTTTTGA CAGATTGTTG AAATTAAATG AATTGAAAGG GAACTCAGA GTAGTAGGAC	360
	GTTTATTAAA AGGAAAAAA TGCTTGCAA TGTGCTGTAA TCACAAGAGG AGAAAAATAC	420
	TTGTTTCCTT GATCTGTGAG AGGTCACAGT AACCTGGGCC GAGCTGTTAT TATTTATTAT	480
10	ATAATAGTAG TAGGAAGTTA ATAACCTGGT CTCTGTGTTT CAAGCACAAT ATTACAACCT	540
	CTTTTGAACC GTAAATATCA GAATGAATCC TCTTCCCAGG GGATTGAACA GAAGCTTAAT	600
15	GTTTACAAGT GTTTGAATTT GTGATCTGAA ATAACACAAA ATTAAAAACA TGATTTCTCT	660
	AATTTTCCAA CTAGAGGAAG AGAACTTGT GGAAAAGTTC TTTTMTTTC TTTTMTTTC	720
	CTTAAAGAAG GGCAGCCAAG GTAGTAACCT AAAAATAGTG CCCAGGCATA TGAGAGTTGT	780
20	CCTACGAGGT TAAAGAACAC ACTGTTCCAC TGTATGGCTT TGGCCCTGAG TGGCCAGGGA	840
	GGTCAACTTG ACCCTGCCAT GTTGGTTTGA CTTACTAAGA CACAGGAATC ATTGTTTTC	900
25	TTGACCAGGG TCTCACACC TGGAGGAATG TTAAGTAAGA GAAAGAACCT CTTTCTGAA	960
	TATTGACATG TAAAGACCA AAGTAATTTT TCTGAACCTC TGCAATTCTG AGAACTCTCC	1020
	AAGGAATTTA CAGTGATTTT AGTGCTTGTG AGCATTTTTC CATGAGGACT TTCATACATT	1080
30	TGACTCTTTA GTTCACAGGT TCCCATGTAT TGTGAGCAAG ATATTTATCT CTTAGCCCT	1140
	TGGGGATCCA GCTGAGAGCA ATCTCTTGCA TTTTMTTACC CGTGTATGTA CAGATATCAT	1200
35	TTCTTGTTGA TGCCATGACT TGAAAAAGTT TGGGAAGCTC TTTAGCAATA TCAGCTAAAA	1260
	GGATATGAAA TCACAGGTGA TAGCAGTTGT CATTAGTAA TTTCTACAA GCAGCACCCC	1320
	AAAGGAAATA TAGTCTAAT CTTTACTATC CACTTCTAAA TTTAATGTGA ATTTCATACA	1380
40	TGTTATTAGT TGTTTCTTT ATAATTTTAT AAAAATTATT CATCGGGAGT TTAACCTCCA	1440
	CTTCCATGCT ATCGGATGTG TTGGGCTCCA TGCAAGAACT TGGAAGAAAA ACAGGCAGGA	1500
45	ATGCATTTGC ATAATGACCC AGATCATCAT TTTCTGCAAC TGAGAATTAT ATTTATCAT	1560
	TGCTTCTAGA AGTCTGCAAT TCTTTACTTT TCTTTGGTGC ATTATTATCT AGGTGCCATC	1620
	ACTGGATAAT GTGGAGTGAC TAGAGAAGTC AYATATCACT GTAAGGTACA GTTAGGGTA	1680
50	ACACTTTAGA GGTATTATT TTTTAAAAA CTTTCTTGA ACTCCTGGGC CAACATGGGT	1740
	GAAACCCCGT CTTCTTACTT AAAAATACCC AAAATTAGGC CAGGGCCGTG GATGGGTGGG	1800
55	GTGCTGTGTA ATCTTCAGCT ACTTNGGGGA GGGCTTGAAG CCAGGGAGGA ACTGCCCTGG	1860
	ANCCCCGGGG NGGGCCAGNA GGTTCGCCAG TTGAGT	1896

(2) INFORMATION FOR SEQ ID NO: 56:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1753 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

5	TCTTTTAAAA ATAGACATTT GTGGGCTCA CACAATATAT GAAATAGTAC CCTCTAAAAA	60
	AGAGAAAAAA AAAATCAGGC GGTCAAACCT AGAGCAACAT TGTCTTATTA AAGCATAGTT	120
15	TATTTCACTA GAAAAAATTT AATATCAAGG ACTATTACAT ACTTCATTAC TAGGAAGTTC	180
	TTTTTAAAT GACACTTAAA ACAATCACTG AAAACTTGAT CCACATCACA CCCTGTTTAT	240
20	TTTCCTTAAA CATCTTGAA GCCTAAGCTT CTGAGAATCA TGTGGCAAGT GTGATGGGCA	300
	GTAAAATACC AGAGAAGATG TTTAGTAGCA ATTAAAGGCT GTTTCACCT TTAAGGACCA	360
	GCTGGGCTGT AGTGATTCCT GGGGCCAGAG TGGCATTTATG TTTTACAAA ATAATGACAT	420
25	ATGTCACATG TTTGCATGTT TGTTCGCTG TTGAATTTTT GAACAGCCAG TTGACCAATC	480
	ATAGAAAGTA TTACTTTCCT TCATATGGTT TTTGGTTCAC TGGCTTAAGA GGTTCCTCAG	540
	AATATCTATG GCCACAGCAG CATACCAGTT TCCATCCTAA TAGGAATGAA ATTAATTTTG	600
30	TATCTACTGA TAACAGAATC TGGGTCACAT GAAAAAAAT CATTTTATCC GTCTTTAAG	660
	TATATGTTTA AAATAATAAT TTATGTGTCT GCATATTGCA GAACAGCTCT GAGAGCAACA	720
35	GTTTCCCATT AACTCTTCT GACCAATAGT GCTGGCACCG TTGCTTCCTC TTTGGGAAGA	780
	GGAAAGGGTG TGTGAACATG GCTAACAATC TTCAAATACC CAAATTGTGA TAGCATAAAT	840
40	AAAGTATTTA TTTTATGCCT CAGTATATTA TTATTTAATT TTTAGGTAA TGCCTATCTC	900
	TTGGTCTATT AAGGAAAGAA GCAATCAGTA GAGAATTCAG GATAGTTTGT TTTAAATTCT	960
	TGCAGATTAC ATGTTTTTAC AGTGGCCTGC TATTGAGGAA AGGTATTCTT CYATACAACT	1020
45	TGTTTTAACC TTTGAGAACA TTGACAGAAA TTATGCAATG GTTTGTTGAG ATACGGACTT	1080
	GATGGTGCTG TTTAATCAGT TTGCTTCCAA AGTGGCCTAC TCAAGAGGCC CTAAGACTGG	1140
50	TAGAAATTAA AAGGATTTC AAACTTTCT ATTCTTTCT TAAACCTACC AGCAAACCTAG	1200
	GATTGTGATA GCAATGAATG GTATGATGAA GAAAGTTTGA CCAAATTTGT TTTTTTGTG	1260
	TTGTTGTGTG TTTGAATTTG AAATCATTCT TATTCCCTTT AAGAATGTTT ATGTATGAGT	1320
55	GTGAAGATGC TAGCGAACCT ATGCTCAGAT ATTCATCGTA AGTCTCCCTT CACCTGTTAC	1380
	AGAGTTTCAG ATCGGTCCT GATAGTATGT ATTTCTTTAG TAAGAATGTG TTAAAATTAC	1440
60	AATGATCTTT TAAAAAGATG ATGCAGTTCT GTATTTATTG TGCTGTGTCT GGTCTAAGT	1500

GGAGCCAATT AAACAAGTTT CATATGTATT TTTCCAGTGT TGAATCTCAC AACTGTACT 1560  
TTGAAAATTT CCTTCCATCC TGAATAACGA ATAGAAGAGG CCATATATAT TGCCTCCTTA 1620  
5 TCCTTGAGAT TTCCTACCT TTATGTTAAA AGTTGTGTAT AATTGTTAAA ATCTGTGAAA 1680  
GAATAAAAAG TGGATTTAAA TTAATAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1740  
AAAAAAAAAGG GGG 1753  
10

## (2) INFORMATION FOR SEQ ID NO: 57:

15

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1220 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

20

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

CGGGAAGTTA CTGCAGCCGC GGTGTTGTGC TGTGGGAAG GGAGAAGGAT TTGTAAACCC 60  
25 CGGAGCGAGG TTCTGCTTAC CCGAGGCCGC TGCTGTGCGG AGACCCCGG GTGAAGCCAC 120  
CGTCATCATG TCTGACCAGG AGGCAAAACC TTCAACTGAG GACTTGGGGG ATAAGAAGGA 180  
30 AGGTGAATAT ATTAACTCA AAGTCATTGG ACAGGATAGC AGTGAGATTC ACTTCAAAGT 240  
GAAATGACA ACACATCTCA AGAACTCAA AGAATCATAC TGTCAAAGAC AGGGTGTTC 300  
AATGAATCA CTCAGTTTC TCTTTGAGG TCAGAGAATT GCTGATAATC ATACTCCAAA 360  
35 AGAACTGGGA ATGGAGGAAG AAGATGTGAT TGAAGTTTAT CAGGAACAAA CGGGGGGTCA 420  
TTCAACAGTT TAGATATTCT TTTTATTTTT TTTCTTTTCC CTCAATCCTT TTTTATTTTT 480  
40 AAAAATAGTT CTTTGTAAAT GTGGTGTICA AAACGGAATT GAAACTGGC ACCCATCTC 540  
TTTGAAACAT CTGGTAATTT GAATCTAGT GCTCATTATT CATTATTGTT TGTMTTCATT 600  
GTGCTGATTT TTGGTGATCA AGCCTCAGTC CCCTTCATAT TACCCTCTCC TTTTAAAAA 660  
45 TTACGTGTGC ACAGAGAGGT CACCTTTTTC AGGACATTGC ATTTTCAGGC TTGTGGTGAT 720  
AAATAAGATC GACCAATGCA AGTGTTCATA ATGACTTTCC AATTGGCCCT GATGTTCTAG 780  
50 CATGTGATTA CTTCACTCCT GGACTGTGAC TTTCACTGGG AGATGGAAGT TTTTCAGAGA 840  
ACTGAAGTGT GGAAAAATGA CCTTTCCTTA ACTTGAAGCT ACTTTTAAAA TTTGAGGGTC 900  
TGGACCAAAA GAAGAGGAAT ATCAGGTGTA AGTCAAGATG ACAGATAAGG TGAGAGTAAT 960  
55 GACTAACTCC AAAGATGGCT TCACTGAAGA AAAGGCATTT TAAGATTTTT TAAAAATCTT 1020  
GTCAGAAGAT CCCAGAAAAG TTCTAATTTT CATTAGCAAT TAATAAGCT ATACATGCAG 1080  
60 AAATGAATAC AACAGAACAC TGCTCTTTTT GATTTTATTT GTACTTTTTG GCCTGGGATA 1140

TGGGTTTTTAA ATGGACATTG TCTGTACCAG CTTTCATTAAA ATAAACAATA TTTGTAAAAA 1200

TCAWAAAAAA AAAAAAAAAA 1220

5

(2) INFORMATION FOR SEQ ID NO: 58:

10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1049 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

15

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

20 TCGCGCCTGC AGACACAGCA TCTACTCAGC GTGGGTACC TCTGTGAACA TCACTGACTG 60

CAAGCCTCCC TCAATTTCTG GTGCAGCCCA TCAGGGACCC ACAGCGCCTG GGAGGATGGT 120

GCGGATCTTG GCCAATGGGG AAATCGTGCA GGACGACGAC CCCCAGTGA GGACCACTAC 180

25 CCAGCCACCA AGAGGTAGCA TTCCTCGACA GAGCTTCTTC AATAGGGGCC ATGGTGCTCC 240

CCCAGGGGGT CCTGGCCCCC GCCAGCAGCA GGCAGGTGCC AGGCTGGGTG CTGCTCAGTC 300

30 CCCCTTCAAT GACCTCAACC GGCAGCTGGT GAACATGGGC TTTCCGCAGT GGCATCTCGG 360

CAACCATGCT GTGGAGCCGG TGACCTCCAT CCTGCTCCTC TTCTGCTCA TGATGCTTGG 420

TGTCGTGGC CTCCTCCTGG TTGGCCTTGT CTACCTGGTG TCCACCTGA GTCAGCGGTG 480

35 ACCTCTGAGG GCTGATAGGG GTGGGTTTGT TGAGAGGGAC TTGCTGGGCC TTGGTGTGAG 540

AGCAGGCATA TTTGGAGGGG ATCTGGTGGT GCCTTGAAGG TATGATCAGA GAGGGGACCA 600

40 CAGGTGTGTG TTTCCCCCTT GTGTTAAGCG TGAGGCAGAG GGAGACGTTA GTCCCAGCAT 660

TTCCCAAAGT GTGGGTGGGT CCGTTGGTTC CCGAGATACT TTTAGGTGGT ATGGGGCCTG 720

CATTAAGTGG CACAAATCA GAGCAAGAAA GCGATGCCCT TCCCAATTCT CTCAATCCTT 780

45 TTATGCCGAG AAGATCTCAG CTGGATGCCA ACATGTTCCG ATGCCTGTGG AAGACATGCC 840

GACGTCTCCT CTGCCTAGGG AGCAGGACTT GGGCTTAGGG CAGGTGGAAA AAATCCAGA 900

50 CTTTTTTAGC ACTGTTTTTG TTTTAATGGT ATATTTTTAT TGGCTACTTT ATTGTTTAGG 960

ACAAGTGGTA GTGGCATTCT ATTTATGTG ACCTTTTCAA TAAATAGATT TAAGTAAAAA 1020

AAAAAAAAAA AAAACTCGAG GGGGGCCCC 1049

55

(2) INFORMATION FOR SEQ ID NO: 59:

60

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1776 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

	AAAGAGGATG TGMAGCTAGA GGTCCCCGAT GGCTGGTCGG ATGGGAAGCA CAAGGCTGAG	60
10	GGACTGGATT GTAAAGGCAC TAAGTCGTTT TCGGTGAGA ATCAGACATG GGGGACCTCT	120
	AGCTTCACAT CCTCTTTCCT TGCAGSTCTG GACATCCTGA GCCCAAGTCC CCCACACTCA	180
	GTGCAGTGAT GAGTGGCGAA GTGAAGGTGA CAGGGCAGAA CCAGGAGCAA TTTCTGCTCC	240
15	TAGCCAAGTC GGCCAAGGGG GCAGCGCTGG CCACACTCAT CCATCAGGTG CTGGAGGCCC	300
	CTGGTGCTTA CGTGTGTTGA GAACTGCTGG ACATGCCCAA TGTTAGAGAG CTGGCTGAGA	360
20	GTGACTTTGC CTCTACCTTC CGGCTGCTCA CAGTGTGTTG TTATGGGACA TACGCTGACT	420
	ACTTAGCTGA AGCCCGGAAT CTTCCTCCAC TAACAGAGGC TCAGAAGAAT AAGCTTCGAC	480
	ACCTCTCAGT TGTCAACCCTG GCTGCTAAAG TAAAGTGAT CCCATATGCA GTGTTGCTGG	540
25	AGGCTCTTGC CCTGCGTAAT GTGCGGCAGC TGGGAAGACCT TGTGATTGAG GCTGTGTATG	600
	CTGACGTGCT TCGTGGCTCC CTGGACCAGC GCAACCAGCG GCTCGAGGTT GACTACAGCA	660
30	TCGGGCGGGA CATCCAGCGC CAGGACCTCA GTGCCATTGC CCGAACCCTK AANAAAAACC	720
	ATTAAAGTTA CGACGGCAGC AGCAGCGCA GCCACATCTC AGGACCCTGA GCAACACCTG	780
	ACTGAGCTGA GGGAAACCAGC TCCTGGCACC AACCAGCGCC ASCCAGCAAG AAAGCCTCAA	840
35	AGGGCAAGGG GCTCCGAGGG ANCGCCAAGA TTTGGTCCAA GTCGAATTGA AAGRACTGTC	900
	GTTTCCTCCC TGGGGATGTG GGGTCCCAGC TGCTGCTCTG CCTCTTAGGA GTCCTCAGAG	960
40	AGCCTTCTGT GCCCTGGGCC AGCTGATAAT CCTAGGTTCA TGACCCTTCA CCTCCCCTAA	1020
	CCCCAACAT AGATCACACC TTCTCTAGGG AGGAGKCAA TGTAGGTCAT GTTTTGTGTG	1080
	GTACTTTCTG TTTTGTGTA CTTCAATGTT TCCATTGCTC CCCGCTGCCA TGCTCTCTCC	1140
45	CTGTGTTTCT TAAGAGCTCA GCATCTGTCC CTGTTCAATTA CATGTCATTG AGTAGGTGGG	1200
	TAGCCCTGAT GGGGTCGCT CTGTCTGGAG CATAACCCAC AGGCGTTTTT TCTGCCACCC	1260
50	CATCCCTGCA TGCTGATCC CCAGTTCCTA TACCCTACCC CTGACCTATT GAGCAGCCTC	1320
	TGAAGAGCCA TAGGGCCCCC ACCTTTACTC ACACCCTGAG AATTCTGGGA GCCAGTCTGC	1380
	CATGCCAGGA GTCAGTGGAC ATGTTTCATCC TAGAATCCTG TCACACTACA GTCATTTCTT	1440
55	TTCTCTCTC TGGCCCTTGG GTCCTGGGAA TGCTGCTGCT TCAACCCAG AGCCTAAGAA	1500
	TGGCAGCCGT TTCTTAACAT GTTGAGAGAT GATTCTTTCT TGGCCCTGGC CATCTCGGGA	1560
60	AGCTTGATGG CAATCCTGGA AGGGTTTAAT CTCCTTTTGT GAGTTTGGTG GGAAGGGAA	1620

GGGTATATAG ATTGTATPAA AAAAAAAG GTATATATGC ATATATCTAT ATATAATATG 1680  
ACGCAGAAAT AAATCTATGA GAAATCTATC TACAAAMWAA AAAAAAAAAA AAAAAAAAAA 1740  
5 AGGAATTCGA TNTCAAGCTT ATCGATACCG TCNACC 1776

10

(2) INFORMATION FOR SEQ ID NO: 60:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 443 base pairs  
15 (B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

ACAGATAAAT AAATAAATAA TAAATTAAAT TAAATAAAAA ATCTGAGCTA ATCTGAATAA 60  
ATTGAGAGAT TTCACATGAA AGCCAGGATT TCTGGCTTCC CAGGAACAGT CAGAAGAGCT 120  
25 AGCTAGCAAC ACTGGTCTGC TTGGCTACCT TCTTTGGAAC AACATGAAAT CTAGCTCCCT 180  
TTTTTTTTTT TTTTGGCCC ACTTCATCCA TTCACATGAC CTGCCTGGCC TCTGCAGGTA 240  
AGTGAGTATG CAACAAAAAT GTAGCACAGG TTTTGTGCTT GAACTACGTG GTTTCAGGTC 300  
30 CAGCTCTGCC ACTTGCTAGC ATGACCTCGT GCCGAATTCC NGCACGAAGT TTTTTTTTTT 360  
TTTTTCAGTG CTCCAGTCCC CCTATTGGAG AATCCTGCCC CCCCTGGGA CAGAATGTTT 420  
35 ACCCTGGCCC CGCGANTCCC TGA 443

40

(2) INFORMATION FOR SEQ ID NO: 61:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2888 base pairs  
(B) TYPE: nucleic acid  
45 (C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

TTAATGTTGT CAATAACCAC CAGGCCAAAC AGAATTTATA TGACCTGGAT GAAGATGATG 60  
ATGGTATAGC TTCCGTTCTT ACTAAACAGA TGAAGTTTGC AGCCTCAGGC GNCCTTCTCC 120  
ACCACATGGC TGGGCTAAGC AGTTCCAAGC TTTCCATGTC CAAGGCCCTC CCTCTCACCA 180  
55 AAGTGGTCA GAATGATGCA TACACAGCTC CTGCTCTCCC TTCTCTATT CGAACAAAAG 240  
CCTTGACCAA CATGTCCCGG AACTGGTGA ACAAGGAAGA ACCCCCCAAA GAGCTGCCAG 300  
60 CTGCTGAGCC TGTCTCAGC CCATTGGAAG GCACCAAGAT GACTGTGAAT AATCTGCACC 360



	CTCGAGTCAC TGAGGAGGAC ATTGTTGAGC TTTCTGTGT GTGTGGGGCC CTCAAGCGAG	420
5	CTCGACTGGT CCATCCTGGG GTAGCGGAGG TGGTGTTTGT GAAAAAGGAC GATGCCATCA	480
	CCGCATATAA GAAGTACAAC AACCGGTGTC TGGACGGGCA GCCGATGAAG TGCAACCTTC	540
	ACATGAATGG GAATGTTATC ACCTCAGACC AGCCCATCCT GCTGCGGCTG AGTGACAGCC	600
10	CATCAATGAA AAAGGAGAGC GAGCTGCCTC GCAGGGTGAA CTCTGCCTCC TCCTCCAACC	660
	CCCCTGCGA AGTGGACCCT GACACCATCC TGAAGGCACT CTTCAAGTCC TCAGGGGCCT	720
15	CTKTGACCAC GCAGCCCA GAATTCAAAA TCAAGCTTTG AGCAGGGGAG TGAGGCAGCC	780
	AGAAGTGGGG GCAGAGGAGG GTGGCTCTGT TTCCCCAAGG CAAAGCTTAT GACCAATGGG	840
	CCATCGGACT GGAGACCCCT GATTGTGGGA AGGGTTGCCA GGGATAAAGA GCTTCCTCAC	900
20	TGGATGGGAC CCGCCTTTCT GTGTGTGTGT CTGCCCTGTG CTCTCTCTC TACGTTAAG	960
	TTTCCTGTAG TATGTTTCTT CATCTCATCG CCAAGGTAGG CTTGTGTMTT TCAGTGTGTG	1020
25	CCTCCCGAG CCTCAGCCCC AAGCTGATTT CTTATCTGGA AATGGTACAC TGAATTCTCT	1080
	GGGTGGCTTT CTGTGGGCC CATGGGATGC AGCGTGGGGG CTGTCTGAAG GACCCCTGCTT	1140
	TTTCCAGGGG CCGAGGGGCT GCCTTCCTT TGTGTGTATT AAGCTTTTCA AACAAATGGAG	1200
30	GGGATGGAGA GCCCTGGTGT CCTGACGGGA GCCAGGTGG CCTGAGAGCT GTGCCGCTCC	1260
	TCTGTCTTGT CAGTGGAGGT GCCTGGGTGG GGAGCAGGTC TCAGGCCTCT TGTCTCTCC	1320
35	CCAGTGGCTC CAGGCCTCAC TAGTGGCAAG GGCAGGATGA GGCTGCACCG CTGGGAAGAG	1380
	TCTATCTAAG YTCITGGCTT GGAGTCCCGT GTGCTCTCCR CCCAGAGGAA GTTCTCCAGA	1440
	GTTACCTTT CCTTTTCCT TGAGTTGTGC TGAATGCCCC ACCCCAGCTC TCTTTCCCTT	1500
40	CTGGGTGTCT TTGCTGGGAG GGGCTGTGT TGTGAGCCCT CCCGTTCTC ACCTCGCCTG	1560
	GCACPTAACC ACACCCCTGGT TTTGTGTAGC CGCCAGCTCT CTTCTGGTGT GGCCPTTGAA	1620
45	AGGCTCAGCC TCCCATTTGT CAGTGTCTGG GTTTGGAGCT TATTTGAATG GAAGAGGTCA	1680
	GTTTGTTCCT GGCTCTCCAT TTCTGGCCTC AGTTGTCTAC AGGACAGTGG TCAGGGATGC	1740
	CTGGAGGCAT ATATCCAGCT GCCACCAAGG GGCAGTGTTC GTTCCACIT ATGTGAGTGA	1800
50	CCCCATCCAT CCATGACCAG AGGATTATTT TCCTGCCTTG GCAGAGGAGG AGGAGTCAAG	1860
	GGAGCAGGGC AGCTCTACCA GGCAAGGTGT TTCCCCAGCA TAGGCGCAGA CAGTTGGGAC	1920
55	GAAACTTCAG AGCCCAGGCA GTCCCTGAAT GACCAGGCCA GTGTGTCTAC TGAGTGGTCC	1980
	CCTGCTGGTT GGGAGTGAAG AGAATCCAGG CTGGCAGAGC TGGAGCCAGT TGGGGAGCAC	2040
	GGTTCGGGA GCTCTGCAAA ATCAGTAGCA AGTGCTGGAA AAGGCACATG CCGAAGATAC	2100
60	TCAAGAGCTC CCAAGATTTG CTTGAGGCTA GCCAGTGAA RAAAACCAGA GACTCATGTT	2160

	TCCAGGGGTC AGTCTGTGAG GCAGGAAGGA CCCAGGATTT GAACCCAGCT TCAGTGTGCA	2220
5	GGCTCTGAGG CTGCCCAGGA CGGAAAGTC CAAGGAAGGG GCCTGGTGGT GCTCCACTTG	2280
	CAGTTCTTTA AAGAATGCTG CTTTTTATTC TCCTAACCTT TTCAAGTGGG TGCAGACTTC	2340
	TCGTTAGCAG CTGGAAGACA TTCCTCCAC ACTTTTCCCT TCCTGGCCCA AGAGAGCATC	2400
10	CAGAAGGCAG TAGGACCTGG TTTTTCAGGT ACTGGGAGCC GGGGGCTCAC TGCTTGCACT	2460
	GTGCTTAGGG TAGGGATGGT AAATATCCTC CCTGCATGGC TTTATCCTCC CTCTCATCCC	2520
15	AAAGCAGGTA TCTTCTGGTT GTCACAGAGT TTCATTGAGT CCAGCTGCAG CCACGTGGCC	2580
	ATCTGGAGCT GGTGCTATAG GTGACCATCT GTTACATTGA GGGGACCTGT TTGCTCTCTC	2640
	CACTCTATAA GCAGTCATCT TGGGAGACCG GGAGGAGAAG GTGGTGGGCT AGTCTCTGTG	2700
20	CCTCTCCAC TTCCCATGCC TCTATGTTAC CCATCTGTGT CTCTGTGCA GAAGGAGAGG	2760
	AAGGGCAATT AAGAGATGAA GGGTGATTAT GTATTACTTA TCCATTTCTG AATAAACATT	2820
25	TGTTATTCCT AAAAAAAAAA AAAAAAACT CGAGGGGGGG CCCGGWACCC AWATCGCCSK	2880
	AAAGTGAG	2888

30

(2) INFORMATION FOR SEQ ID NO: 62:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1851 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

35

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

40	CACTAGTATA ATTTATAATT ATAACTATT CTGATTCTT TTCAAATATT AGGTGTCCTA	60
	GTGCTATATG AAGGTTTGCC ACTTCATCTT GCACTGTTCC CCAAACCTTG GACTGAGCTA	120
45	TGCCAGACTC AGTCTGCTAT GTCAAAAAAC TGCATCAAGC TTTTGTGTGA AGATCCTGTT	180
	TTCCGAGAAT ATATTAAATG TATCCTAATG GATGAAAGAA CTTTTTTAAA CAACAACATT	240
50	GTCTACACGT TCATGACACA TTTCTTCTA AAGGTTCAAA GTCAAGTGTT TTCTGAAGCA	300
	AACTGTGCCA ATTTGATCAG CACTCTTATT ACAAACCTGA TAAGCCAGTA TCAGAACCTA	360
	CAGTCTGATT TCTCCAACCG AGTTGAAATT TCCAAAGCAA GTGCTTCTTT AAATGGGGAC	420
55	CTGAGGGCAC TCGCTTTGCT CCTGTCAGTA CACTCTCCA AACAGTTAAA CCCAGCTCTA	480
	ATTCCAACCTC TGCAAGAGCT TTTAAGCAAA TGCAGGACTT GTCTGCAACA GAGAAACTCA	540
60	CTCCAAGAGC AAGAAGCCAA AGAAAGAAAA ACTAAAGATG ATGAAGGAGC AACTCCCAT	600

	AAAAGGCGGC GTGTTAGCAG TGATGAGGAG CACACTGTAG ACAGCTGCAT CAGTGACATG	660
	AAAACAGAAA CCAGGAGGT CCTGACCCCA ACGAGCACTT CTGACAATGA GACCAGAGAC	720
5	TCCTCAATTA TTGATCCAGG AACTGAGCAA GATCTTCCTT CCCCTGAAAA TAGTTCTGTT	780
	AAAGAATACC GAATGGAAGT TCCATCTTCG TTTTCAGAAG ACATGTCAAA TATCAGGTCA	840
10	CAGCATGCAG AAGAACAGTC CAACAATGGT AGATATGACG ATTGTAAAGA ATTTAAAGAC	900
	CTCCACTGTT CCAAGGATTC TACCCTAGCC GAGGAAGAAT CTGAGTTCCC TTCTACTTCT	960
	ATCTCTGCAG TTCTGTCTGA CTTAGCTGAC TTGAGAAGCT GTGATGGCCA AGCTTTGCCC	1020
15	TCCCAGGACC CTGAGGTGCG TTTATCTCTC AGTTGTGGCC ATTCCAGAGG ACTCTTTAGT	1080
	CATATGCAGC AACATGACAT TTTAGATACC CTGTGTAGGA CCATGGAATC TACAATCCAT	1140
20	GTCGTCACAA GGATATCTGG CAAAGGAAAC CAAGCTGCTT CTTGACATTA GGTGTAGCAT	1200
	GTCTACTTTT AAGTCCCTCA CCCCCAACCC CCATGCTGTT TGTATAAGTT TTGCTTATTT	1260
	GTTTTGTGTC TTCAGTTTGT CCAGTGCTCT CTGCTTGAAT GGCAAGATAG ATTTATAGGC	1320
25	TTAATTCCTG GTCAGGCAGA ACTCCAGATG AAAAAAAGT GCATCTTCAG TATACTTCCT	1380
	AAAGGGCAAT CAGATAATGG ATATGTTTTA TGTAAITTAAG AGTTCACITTT AGTGGCTTTC	1440
30	ATTTAATATG GCTGTCTGGG AAGAACAGGG TTGCCTAGCC CTGTACAATG TAATTTAAAC	1500
	TTACAGCATT TTTACTGTGT ATGATATGCT GTCTCTGTG CCAGTTTGT ACCTTATAGA	1560
	GGCAGATTGC CTCGATCGC TGTGGTCTT ATATCAAAA TTAAGTTTAC TTGTATACGG	1620
35	AACAACCACA AGAAATTGA TTCTGTAAAG AATCCTCTTT AGCTGTGGCC TGGCAGTATA	1680
	TAAATGGTGC TTTATTTAAC AGAATACCTG TGGAGGAAAT AAAGCACACT TGATGTAAAA	1740
40	ATAATGTGTT TATTTTATT GACATGACTG ATTGATGCT ATTCTGTGCA CTTAATTAAA	1800
	CTGATTGTGA TGACTTWWAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA A	1851

45

(2) INFORMATION FOR SEQ ID NO: 63:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3542 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

50

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

55

	TCCAATGCTG ATGAGCGTCT TCGCTGCCAG GCCAGCTCCT TGCCTGCTGA TGACCTTTGC	60
	ACAGAAAATG CCATCATGCT GAAACGATTC AATAGGTATC CGCTGATCAT TGACCCCTCT	120
60	GGACAGGCCA CAGAATTCAT TATGAATGAA TATAAGGWTG GTAAGATCAC ACGGACCAGC	180

	TTCTGGATG ACGCCTTCAG AAAGAACTTA GAGAGTGCAC TGAGATTCGG TAACCCCTT	240
5	CTGGTCCAGG ATGTGGAAAG CTACGATCCA GTTTTGAACC CGGTGCTGAA CCGTGAAGTG	300
	CGGCGAACAG GGGGAGAGT GCTGATCACT CTCGGGGACC AGGACATAGA CCTGTCGCCA	360
	TCGTTTGTC TCTTCTGTG CACCCGGGAT CCAACTGTG AGTTCCACC AGATCTCTGT	420
10	TCCCGGGTTA CTTTGTAA CTTACAGTT ACCCGTAGCA GTTTACAAAG CCAGTGTCTA	480
	AATGAAGTAC TTAAAGCAGA AAGACCTGAT GTGGACGAGA AACGATCTGA TCTTCTTAA	540
15	CTTCAAGGG AATTTCAGCT CGTTTTCGT CAGCTGAAA AATCTCTACT ACAAGCTCTG	600
	AACGAGGTGA AAGGGCGCAT TTTGGATGAC GACACGATCA TAACCACTCT GGAGAACCTG	660
	AAGAGAGAGG CTGCAGAGT CACCAGGAAA GTTGAGGAGA CGGACATTGT CATGCAGGAG	720
20	GTGGAGACCG TGTCCAGCA GTACCTCCG CTCTCCACCG CCTGCAGCAG CATCTACTTC	780
	ACCATGGAGT CCCTCAAGCA GATACACTTC TTGTACCAGT ACTCCCTCCA GTTTTTCCTG	840
25	GACATTTATC ACAACGTCCT ATACGAGAAC CCGAACCTGA AGGGTGTAC CGACCACACA	900
	CAGCGCTGT CCATTATAAC AAAGGACCTC TTCCAGGTGG CGTTTAACCG AGTGGCTCGA	960
	GGCATGCTGC ATCAGGACCA CATTACCTTT GCCATGCTGC TGGCAAGAAT CAACTGAAG	1020
30	GGCACCCTGG GGGAGCCAC CTACGATGCA GAATTCAGC ACTTCTTGAG AGGAAATGAG	1080
	ATTGTCTGA GTGCTGGCTC CACCCAGG ATCCAGGCC TGACTGTGA GCAGGCGGAG	1140
35	GCGGTGGTGA GGCTGAGCTG CCTTCCCGG TTAAAGGACT TGATGCAAA GGTTCAGCA	1200
	GACGAGCAAT TTGGCATCTG GCTGGACAGC AGCTCCCCG AGCAGACTGT GCCCTACCTC	1260
	TGGAGTGAAG AAACACCTGC AACACCCATT GGCCAGGCCA TCCACCGCT GCTCCTGATC	1320
40	CAGGCTTTCC GGCCGATCG CCTGTGGCC ATGGCCACA TGTGTGTTT AACAAACCTT	1380
	GGGAGTCTT TCATGTCCAT CATGGAGCAG CCGCTCGACC TGACCCACAT TGTGGSCACA	1440
45	GAGGTGAAGC CCAACACTCC TGTCTTAATG TGCTCTGTGC CTGGTTATGA TGCCAGTGA	1500
	CATGTGAGG ACCTTGACG CGAGCAGAAC ACGCAGATCA CTTCAATTGC AATCGGCTCT	1560
	GCAGAAGGCT TTAACCAAGC AGATAAGGCA ATAAACACCG CTGTAAAGTC GGCAGGTGG	1620
50	GTGATGCTGA AGAATGTGCA TCTGGCCCCA GGTGGCTGA TGCAGCTGGA GAAGAAGTTG	1680
	CATTCCCTGC AGCCGATGC CTGCTCCGA CTCTCTCA CCATGGAGAT CAACCCAAG	1740
55	GTGCCTGTGA ATCTGCTCCG TCGGGCCGC ATCTTTGTGT TCGAGCCACC GCCAGGKTG	1800
	AAGGCCAACA TGCTGAGGAC GTTCAGCAGC ATTCCCGTCT CACGGATATG CAAGTCTCCC	1860
	AACGAGCGTG CCCGCTGTGA CTTCTGCTG GCCTGGTTTC ATGCGATCAT CCAAGAACGC	1920
60	TTACGATACG CACCACTGGG GTGGTCAAAG AAGTATGAAT TTGAGAGTC TGACCTGCGG	1980

	TCANYTTGCG ATACGGTGGA CACGTGGCTG GATGACACGG CCAAGGGCAG GCAGAACATC	2040
5	TCACCGGATA AGATCCCGTG GTCTGCACTA AAGACCTTAA TGGCCAGTC CATTTATGGC	2100
	GGGCGCGTGG ACAACGAGTT TGACCAGCGT CTGCTCAACA CCTTCCTGGA GCGCCTGTTT	2160
	ACAACCAGGA GTTTCGACAG TGAGTTTAAG CTGGCATGCA AGGTCGACGG ACATAAAGAC	2220
10	ATTCAAATGC CAGATGGCAT GCAGGCGAGA GGAGTTTG TGAGTGGTGG AGTTGCTCCC	2280
	CGACACCCAG ACGCCCTCCT GGCTGGGCCT GCCCAACAAC GCCGAGAGAG TCCTCCTTAC	2340
15	CACACAGGGT GTGGACATGA TCAGTAAAT GCTGAAGATG CAGATGTTGG AGGATGAGGA	2400
	CGACCTGGCC TACGCAGAGA CTGAGAAGAA GACGAGGACA GACTCCAGT CCGACGGGG	2460
	CCCTGCCCTG ATGCGGACAC TGCACACCAC CGCGTCCAAC TGGCTGCACC TCATCCCCCA	2520
20	GACGCTGAGC CACCTCAAGC GCACCGTGGA GAATATCAAG GATCCTTTGT TCAGGTTCTT	2580
	TGAGAGAGAA GTGAAGATGG GCGCAAAGCT GCTTCAGGAC GTTCGCCAGG ACCTTGCGA	2640
25	TGTCGTCCAG GTGTGCGAAG GAAAGAAGAA GCAGACCAAC TACTTGCGCA CGCTGATCAA	2700
	CGAGCTAGTG AAAGGGATCT TGCTCGGAG CTGGTCCAC TACACGGTGC CTGCCGGCAT	2760
	GACCGTCATC CAGTGGGTGT CCGACTTCAG CGAGAGGATC AAACAGCTGC AGAACATCTC	2820
30	ACTGGCAGCT GCATCTGGTG GCGCCAAGGA GCTAAAGAAC ATCCACGTGT GCCTGGGTGG	2880
	CCTGTTCTGT CCTGAGGCGT ACATCACTGC CACCAGGCAG TATGTGGCCC AGGCCAACAG	2940
35	CTGGTCCCTG GAGGAGCTCT GCCTGGAAGT CAACGTCACC ACCTCACAGG GCGCCACCCT	3000
	TGACGCTTGC AGCTTCGGAG TCACGGGTTT GAACTTCAA GGGGCCACGT GCAACAACAA	3060
	CAAGCTGTCA CTGTCCAATG CCATCTCAAC CGCCCTTCCC CTGACGCAGC TGCGCTGGGT	3120
40	CAAGCAGACA AACACCGAGA AGAAGGCCAG TGTGGTAACC TTACCTGTCT ACCTGAACCT	3180
	CACCGTGCA GACCTCATCT TCACCGTGA CTTCGAAATT GCTACAAAGG AGGATCCTCG	3240
45	CAGCTTCTAC GAGCGGGTG TCGAGTCTT GTGCACAGAG TAACTTTTC TAGTGCCCC	3300
	TTTCTGTAAT AGTGAAAGTT GGTATTAAAC ATTTATTCAT TTTTAAATA TTTGGAAGGT	3360
	CTGAGCTTGT GAAAAGAAAG TGGTTGGTCT GAGGTTGGAG GAAGCTGAAT GGAATCTGAC	3420
50	GGTGGGAGT GGTGAAATT GGAAGGATAC CAGGAGGTAT TTGGGAAGGC CAATGGCGTG	3480
	GCTCCTTTGA GGAAATAAAA CACTAAGCAT GAAAAAATAA AAAAACTTA CAANCCNCAA	3540
55	GG	3542

60

(2) INFORMATION FOR SEQ ID NO: 64:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 883 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

5  
10  
15  
20  
25  
30  
35  
40  
AGGTGATTTT AATGATAGGT GTCATATATA GGACGGATAA TCTGTTTACA TTCTGTTCTT 60  
CTCGATGCAC TCACAAGCGG GTAAC TAGGT GACAAGAAAA CAAAGATCTT ATTCAAAGA 120  
GGTCTTACAG CAACCCAACG TCTCATCTTC CCATAGTAAA GATGACGGCG CCTTGAGGTA 180  
AGCTACAGGC AACACCACTT CCGCGTTTCT CTGCGCCCT GGTCCAAGAT GGCGGATGAA 240  
GCCACGCGAC GTGTTGTGTC TGAGATCCCG GTGCTGAAGA CTAACGCGG ACCCGGAGAT 300  
CGTGAGTTGT GGTGCGAGG ACTGAAGGAG GAATATCAGT CCCTTATCCG GTATGTGGAG 360  
AACAACAAGA ATGCTGACAA CGATTGGTTC CGACTGGAGT CCAACAAGGA AGGAACTCGG 420  
TGGTTTGGAA AATGCTGGTA TATCCATGAC CTCCTGAAAT ATGAGTTTGA CATCGAGTTT 480  
GACATTCCTA TCACATATCC TACTACTGCC CCAGAAATG CAGTTCTGA GCTGGATGGA 540  
AAGACAGCAA AGATGTACAG GGTGGCAAA ATATGCCTGA CGGATCATTT CAAACCTTTG 600  
TGGGGCCAGG AATGTGCCCA AATTGGACT AGCTCATCTC ATGGCTCTGG GGCTGGGTCC 660  
ATGGSTGGCA GTGGAATCC CTGATCTGAT TCAGAAGGGC GTCATCCAAC ACAAAGAGAA 720  
ATGCAACCAA TGAAGAATCA AGCCACTGAG GCAGGGCAGA GGGACCTTTG ATAGGCTACG 780  
ATACTAWTTT CCTGTGCATC ACACTTAACT CATCTAACTG TTCCCGGAC ANCCTCCACT 840  
CTAGTTGTTA CTAAGTANTG CAGTAGCATT NTGGGAAGA ACA 883

## (2) INFORMATION FOR SEQ ID NO: 65:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1541 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

50  
55  
60  
GGCACGAGGT GGCCTCTACC CTGGGCTCAT CTGGCTACAC AGGGACTCTA AACGCTTCCA 60  
GATTCCTTGG AAACATGCCA CCCGGCATAG CCTCAACAA GAAGAGGAAA ATACCATTTT 120  
TAAGGCCTGG GCTGTAGAGA CAGGAAGTA CCAGGAAGGG GTGGATGACC CTGACCCAGC 180  
TAAATGGAAG GCCCAGCTGC GCTGTGCTCT CAATAAGAGC AGAGAATTCA ACCTGATGTA 240  
TGATGGCACC AAGGAGGTGC CCATGAACCC AGTGAAGATA TATCAAGTGT GTGACATCCC 300

	TCAGCCCCAG GGCTCGATCA TTAACCCAGG ATCCACAGGG TCTGCTCCCT GGGATGAGAA	360
5	GGATAATGAT GTGGATGAAG AAGATGAGGA AGATGAGCTG GATCAGTCGC AGCACCATGT	420
	TCCCATCCAG GACACCTTCC CCTTCCTGAA CATCAATGGT TCTCCCATGG CGCCAGCCAG	480
	TGTGGGCAAT TGCAGTGTGG GCAACTGCAG CCGGAGGCA GTGTGGCCCA AAACTGAACC	540
10	CCTGGAGATG GAAGTACCCC AGGCACCTAT ACAGCCCTTC TATAGCTCTC CAGAACTGTG	600
	GATCAGCTCT CTCCCAATGA CTGACCTGGA CATCAAGTTT CAGTACCGTG GGAAGGAGTA	660
15	CGGGCAGACC ATGACCGTGA GCAACCTCA GGGCTGCCA CTCTTCTATG GGGACCTGGG	720
	TCCCATGCCT GACCAGGAGG AGCTCTTTGG TCCCGTCAGN CTGGAGCAGG TCAAATTCCT	780
	AGGTCCTGAG CATATTACCA ATGAGAAGCA GAAGCTGTTC ACTAGCAAGC TGCTGGACGT	840
20	CATGGACAGA GGAATGATCC TGGAGGTCAG CGGTCATGCC ATTTATGCCA TCAGGCTGTG	900
	CCAGTGCAAG GTGTACTGGT CTGGGCCATG TGCCCATCA CTGTGTGCTC CCAACCTGAT	960
25	TGAGAGACAA AAGAAGGTCA AGCTATTTTG TCTGGAAACA TTCCTTAGCG ATCTCATTCG	1020
	CCACCAGAAA GGACAGATAG AGAAGCAGCC ACCGTTTGAG ATCTACTTAT GCTTTGGGGA	1080
	AGAATGGCCA GATGGGAAAC CATGGGAAAG GAAACTCATC TTGGTTCAGG TCATTCCAGT	1140
30	AGTGGCTCGG ATGATCTACG AGATGTTTTT TGGTGATTTC ACACGATCCT TTGATAGTGG	1200
	CAGTGTCCGC CTGCAGATCT CAACCCAGCA CATCAAGGAT AACATCGTTG CTCAGCTGAA	1260
35	GCAGCTGPAC CGCATCCTTC AAACCCAGGA GAGCTGGCAG CCCATGCAGC CCACCCCCAG	1320
	CATGCAACTG CCCCCTGCCC TGCCCTCCCA GTAATTGTGA ATGCCATCTT CTTCCCTCTC	1380
	TTTTTTATAA TATTGTACAT ATGGATTTTT TTATTGTTTA GATTTAAACCA GCTTTTAAAT	1440
40	CTCTGTTTTT TGTGACAGTG TTAGAAGTTT GTGATTCTCC AAATATGCCT AGATTTAAAG	1500
	CTGATTTAAT TTATGGAAAA AAAAAAAAAA AAAAAAAAAA A	1541

45

(2) INFORMATION FOR SEQ ID NO: 66:

50

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 732 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

60

AGAAAAATGAA TGTTAGAAGG TGCTGCCGA GCGGGACAG AGTGTTCGCT CGCGCTGGAG	60
AAGGCTCTGC TCAGCCCTGA GAGTCCCTTC CTGCCCCACC GATACTGGCA CTTTAAAAAG	120

	GAAGCTGACC GCACAGTGTG CAGACGAATT GGCCCCAGA AGATGGGGAG TTCTGTCTTG	180
	CCCTTCTGTG TCTGCGTGAC CTCACCCAGC CTAGGAGGGA GGTGCATTCA GGGTAGATTT	240
5	GCCTCTCATT CAAAGTCTG GGGCTTTGGG CGGAAAACAG CCAGCTTTGG CGCTGTGGG	300
	GAGACTCCTC CAGACCAGGA ACCCCAGAAG GAGACAGAGC CTGCCACATC CTCCCACGCC	360
10	AGGCCCTGGG CCAGGGTGAT TGGACTGAGA ATTTGGCCAC AACCAAATTG ATGCTGGCTG	420
	GAACCAGAGG CCAGAAAGCC TGGCCTGTG CCCATGTGGG AGCCCTGTCC TCAGCCCTCT	480
	TGTCCCTTG AGCTCAGTGA ATTCCACCA GGTGCCACA GCTCCTGGAC TTCAAATTCT	540
15	ATATATTGAG AGAGTTGGAG AGTATATCAG AGATATTTT GGAAAGGAGT TGGTCTATGC	600
	AATGTCAGTT TGAATCTTC TTGAAAGTTT AATGTTTTTA TTAGGAGATT TAAAGAAAAT	660
20	AAAGGTCTAC AATATCAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA	720
	AAAAAAAAAA AA	732

25

(2) INFORMATION FOR SEQ ID NO: 67:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 629 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

35

	TTAAGGAATT CGGCMGATC CCGCAAGTA ACATGACTAA AAAGAAGCGG GAGAATCTGG	60
	GCGTCGCTCT AGAGATCGAT GGGCTAGAGG AGAAGCTGTC CCAGTGTCCG AGAGACCTGG	120
40	AGGCCGTGAA CTCCAGACTC CACAGCCGGG AGCTGAGCCC AGAGGCCAGG AGGTCCCTGG	180
	AGAAGGAGAA AAACAGCCTA ATGAACAAAG CCTCCAATA CGAGAAGGAA CTGAAGTTTC	240
45	TTCGGCAAGA GAACCGGAAG AACATGCTGC TCTCTGTGGC CATCTTTATC CTCCTGACGC	300
	TCGTCTATGC CTA CTGGACC ATGTGAGCCT GGCAC TTCCC CACAACCAGC ACAGGCTTCC	360
	ACTTGGCCCC TTGGTCAGGA TCAAGCAGGC ACTTCAAGCC TCAATAGGAC CAAGGTGCTG	420
50	GGGTGTTCCC CTCCCAACCT AGTGTTC AAG CATGGCTTCC TGGCGGCCCA GGCCTTGCTT	480
	CCCTGGCCTG CTGGGGGGTT CCGGTCTCC AGAAGGACAT GGTGCTGGTC CCTCCCTTAG	540
55	CCCAAGGGAG AGGCAATAAA GAACACAAAG CTGAAAAAAA AAAAAAAAAA AACTCGTAGG	600
	GGGGGCCCGT ACCCAATCGC CTTNTCGTG	629

60



## (2) INFORMATION FOR SEQ ID NO: 68:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1751 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60

CTGCTAGCCG GCCGGCGCAG GCTGCCGAGC GGGTGAGCGC GCAGGCCAGG CCAAAGCCCT 60  
GGTACCGCG CGGTGCGGCG CTCAGTCTGC GGCCATGGGG GCGTCCGCGC GGCTGCTGGC 120  
AGCGGTGATC ATGGGGGCCC CGGGCTCGGG CAAGGGCACC GTGTCGTGCG GCATCACTAC 180  
ACACTGCGAG CTGAAGCACC TCTCCAGCGG GGACCTGCTC CGGGACAACA TGCTGCGGGG 240  
CACAGAAATT GCGGTGTTAG CCAAGGCTTT CATTGACCAA GGGAAACTCA TCCCAGATGA 300  
TGTCATGACT CGGCTGGCCC TTCATGAGCT GAAAAATCTC ACCCAGTATA GCTGGCTGTT 360  
GGATGGTTTT CCAAGGACAC TTCCACAGGC AGAAGCCCTA GATAGAGCTT ATCAGATCGA 420  
CACAGTGATT AACCTGAATG TGCCCTTTGA GTTCATTAAA CAACGCCTTA CTGCTCGCTG 480  
GATTTCATCC GCCAGTGGCC GAGTCTATAA CATTGAATTC AACCTCCCA AAAGTGTGGG 540  
CATTGATGAC CTGACTGGGG AGCCTCTCAT TCAGCGTGAG GATGATAAAC CAGAGACGGT 600  
TATCAAGAGA CTAAAGGCTT ATGAAGACCA AACAAAGCCA GTCCTGGAAT ATTACCAGAA 660  
AAAAGGGGTG CTGGAAACAT TCTCCGGAAC AGAAACCAAC AAGATTTGGC CCTATGTATA 720  
TGCTTTCCTA CAAACTAAAG TTCCACAAAG AAGCCAGAAA GCTTCAGTTA CTCCATGAGG 780  
AGAAATGTGT GTAACATATTA ATAGTAAGAT GGGCAAACCT CCTAGTCCTT GCATTTAGAA 840  
GCTGCTTTTC CTAAGACTTC TAGTATGTAT GAATTCCTTG AAAATTATAT TACTTTTATT 900  
TCTACTGATT TTATTTTGA TACTAAGGAT GTGCCAAATG ATTCGGATAC TAAGATGCAT 960  
CGTTTGAAAT CATCTAGTGT GTTGATGCA GTTATCCTCA AAAACATCAG CGATGTCTGA 1020  
ACCTTTAAAA CATCTGTTAG AGCAAAATTA AAAGAGCATT TGGTAGTAAT CTAACCTTTT 1080  
GTTCAAGTAA TAAGTGGTGT ATAAAGTTTC CATATTTTTC TGGAAAAGTT AAAAAAAGTT 1140  
ACATGTCATT TGGAGAAAAT ACGTAATCAG AAATTTGTGC ATAGATTGAT GCCAAAAAG 1200  
ACATTTCCAG CATGTGGAA CATGGTGAGA CACTATATAA AATTCCAGAA AGAAAGCAAC 1260  
TGGATTTACA GATTTATTGT GAGACACAAA TTCACTGCTG CCTTTACACT AAGAAATGTA 1320  
TATGTTAACC ATATATGCTG TATTTATTTT GTGTTAAGC ATACTTTCAG TTTACTCAGA 1380  
ATTTTCAATT TGCTATAAAG ATGTATCAAT TAGCATATAG AAAAATATTA CTTTAAGATG 1440  
ACTTGTTTCC TTTGAAAATA CCTGTGTACT GAGGGTTATG ATTTGTGTCA AAAATTGACA 1500

TAAGTGCTTT TACAAGCACC AAAGTTGAAT GAATTTTCAA CAAAATGTAA TTAAAGTCTA 1560  
TGTTTTTCAGT TATGACTCAG GTTAAGAAAT GTGTTTTAGG ATCTACTTGC TGGTTTTTCT 1620  
5 TTTTGATCCA AATGTGTGAT CTGCCCTGAT AAATAACAAG TTATNGTACC ATCTCCCCCG 1680  
CCAATAAAAA AAAAAAAAAA AAAAAAAAAAC TCGAGGGGGG GCCCGGTACC CAATTCTCCG 1740  
NAATAGGNAG T 1751

10

(2) INFORMATION FOR SEQ ID NO: 69:

15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 508 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

20

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

25

GGCAGGAGAT TATGTATTAA AATGTTTTTG AATTGTGAAA TATTAGAATA TGTACTAT 60

TTGACCCAAC TCAAAATCTC CATGGGAAAA TACCTGTGCA TACCCACAGT ATTGTGAAA 120

ATAATCAGAT GCAGTATCAC AGCTGTGTCA GACTCTAGTA CCAGTTGGGC AATCAAGGCA 180

30

CAGCTAAAAA TTGAAAACAA AGATCTGGAC AACAAAACAG CCAAAGGTGG GGTCAAGAA 240

GCTCTGACGT GTACCTAGCT GTAGAATGCT ATGCACACGT GCCAGGTGTA GTGTGCATAT 300

35

CCAGGAAAAA CTGCAGAGAG CCCAGTCTT CACCTCTGGT TGACCATGAG CTCTGTGTAA 360

GCAGGAAGTG AAGGCTAAGG CAGATTTAAG CTCTGAAAGC ATTCCACAAC ATACACACAA 420

ATCGTGCAAA GCATTAAGGA AATCTTGTTA CTGCTAAGTG TTGCTGACCC AGGAACAAC 480

40

CCTACTCAGC TGGACTTAAA AATAAAAA 508

45

(2) INFORMATION FOR SEQ ID NO: 70:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 245 base pairs

(B) TYPE: nucleic acid

50

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

55

TACATAGAGC AAAGAGAAAT TTCCAGAAAT TCTARAATTC TGGAAAGAGA ATTTTCCTGA 60

GATTGCAGAT TTGCTTGTGT CCTCAGGTGA TGATAGGGG TGTMTTCCCC TGTGTCTCTT 120

60

TCCTCACACT CATGCTTCCT CTCTAGAGT GTCTGGTGG CATGATCATG TGCTACCTAG 180

GCATTTCTTT CACTGATACA AGGAAACTG CAGGTTAAA AAAAAAAAAA AAAAAAAAAA 240  
NCNCG 245

5

(2) INFORMATION FOR SEQ ID NO: 71:

10

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 361 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

ATGTTCTCA TGAGGATGCA CTGTGCTTC TGCAAGTATT GCTGCAGCTT CATAGTGACT 60  
20 CCCACCAGCA CCAGCAATAC AGCTAGCTAC CTGTGGCCTT GGATCTCAGC CAGCATGGCT 120  
GGGAGAGGGA GCAGCTGGGC ATGTACCCTA AATGCTGTTA CCAGGGAAGG ACTCCCAGAG 180  
TGAAGACAAG TAGGGACTTC CTGCAGAGGT GGTACATGTG CTCTCTGTAT CCATACITTT 240  
25 TTTTTTTTTT TTTTGAGATA GAGTTTCACC CTTGTTGCCC TGGCTGGAGT GCAATGGTGC 300  
GATCTCAGCT CACTGCAACC TCTCTGCCTC CCGGGTTCAA GTGATTCTCC TGCCTCAGCC 360  
30 T 361

35

(2) INFORMATION FOR SEQ ID NO: 72:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 713 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

45 AGGATCACAC AATAGAGAAC ACTGTAGTAA CATTTGGGTC TGCTCACAAG ACCCAGAACA 60  
TTGATCAGTT TTTGTTGTTG GTTTATTATT TTCTGTATA AAAATTGTGA AAAGTTGTT 120  
TTAGCTAGAT GATATTTTAA TAGCTGCGAG TGCTTTGGAA CTATAAAGAT GTCACTACTT 180  
50 AACACACATA CCTTATGTTT TGTTTTGTTT TGTTTTACAC TCAGTATAAA TCAGGAGAAG 240  
TTAGCCAACC ATCTAGCATT TAGAATCCTC TTTTATTG TCTTCTAAGG ATATGGATGT 300  
55 TCCCATACA GCAACAAAAC AGCAACAAA ACATTTTATA AATATCACTT GATAGACTGT 360  
AAGCACCTGC TTAACTTTGT GTCCCAAATA TTTAGTGTGT ATATATATAT ATATATATAC 420  
ACACACACAC ACATATATAT TCAACAAATA AAGCAAATA TAACATGCAT TTCACATTTT 480  
60

GTCTTTCCCT GTTACGATT TAATAGCAGA ACTGTATGAC AAGTTTAGGT GATCCTAGCA 540  
TATGTTAAAT TCAAATTAAT GTAAACAGA TTAACAACAA CAAAGAACT GTCTATTGA 600  
5 GTGAAGTCAT GCTTCTATT ATAATAACTT GGCTTCGGTT ATCCATCAA TGCACACTTA 660  
TACTGTTATC TGATTGTTTA TAATAAGAA TACTGTACTT ATAAAAAAAA AAA 713

10

(2) INFORMATION FOR SEQ ID NO: 73:

(i) SEQUENCE CHARACTERISTICS:  
15 (A) LENGTH: 862 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

GAAAGTCAGA GCTGTCCAAT CCCTCAGCAC CTMTTAGATT TGCTCCAAAT TAGAAACGTG 60  
GGGACTATGT GTTCTGGGCA ATCAGAGTC TGGAAATGG CTCTGCAGGC TCTTGATAGT 120  
25 GAGACAGTGG TCATCTTACC AGACATGCAT CTGATTTTAA GCCTCAGGCT AATCCACAAT 180  
GCTCGCCAT GCCTATGATT AACAAACAAA AGCAAAATCT GCTTTTATAG TTTAGGAAAC 240  
30 CTGGATAGAA CAGTATTTT CAGCATTCTT GGATAAGCA GTTCTGCATT TTAAATTGG 300  
GACTGCAGAA GTGACTGTCT ATAGTTGTGA AATACAAAA ATGGTATGTT TGATCAGAAA 360  
AGGAAGCCCG TGCCTGGCAC TTGGAAAGAT ACTGAGCATC ATAACCCTAA TGAGAAAATG 420  
35 TAGGCTCTGT GAATGTTAAC TACAAATCAG GTTAGGAAAG CATATGACAC CCTTGTCAA 480  
ACTAAGCTTC ACTAGGAGGA CCTGTCTCA TAGAAGAATA TGCTTTAAAA GTATCAATTT 540  
40 TCCACAGTCG ATGATGGAGA AAAGTTTATT TGCACCAGAA TGCTGATAGT CACAATACAC 600  
AGCCTGACAT ATATAACAAT ACAGTTTCT GTAAACAGAA GTTCTTCTC TTCCAATTCA 660  
GGAGTCAGTC AGAGCATAAA TATTGCATGT TTCACTTAG AACTGATTC ATTTTAGAAA 720  
45 GCAGATCTGG ATTATTTTGC AGGTAGAAA TGAAGCTAT TTCTGGCATT CTGTCTCAA 780  
AAGTCAATAT ATGTACATTA AGTATAAAAA AGGGTCTCTT TCACCTCTTT TGTTTCGTAG 840  
50 CATTGGCTAC ATAACCTGTG CC 862

55 (2) INFORMATION FOR SEQ ID NO: 74:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 4602 base pairs  
60 (B) TYPE: nucleic acid  
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

5	GCGAGGGGGC GKGGGGAGCA GCGCCGARGC CGCCGCCTCC GCCTCCGCGG CCTAGGACTA	60
	GGGGGTGGGG GACGGACAAG CCCCGATGCC GGGGGAKACG GAAGAGCCGA GACCCCCGGA	120
10	GCAGCAGGAC CAGGAAGGGG GAGAGGCGGC CAAGGCGGCT CCGGAGGACC CGCAACAACG	180
	GCCCCCTGAG GCGGTCGCGG CGGCGCCTGC AGGGACCACT AGCAGCCGCG TGCTGAGGGG	240
	AGGTCCGGAC CGAGGCCGGG CCGCTGCGRC CGCCGCGCMG CAGCTGTGTC CCGCCGGAGA	300
15	AGGCCGAGTA TCCCCGCCGG CGAGGAGCAG CCCAGCGCC AGGCCTCCCG ACGTCCCCGG	360
	GCAGCAGCCC AGGCCCGGAA GTCCCCGTCT CCAGTTCAGG GCAAGAAGAG TCCGCGACTC	420
20	CTATGCATAG AAAAAGTAAC AACTGATAAA GATCCCAAGG AAGAAAAAGA GGAAGAAGAC	480
	GATTCTGCCC TCCCTCAGGA AGTTTCCATT GCTGCATCTA GACCTAGCCG GGGCTGGCGT	540
	AGTAGTAGGA CATCTGTTTC TCGCCATCGT GATACAGAGA ACACCCGAAG CTCTCGGTCC	600
25	AAGACCGGTT CATTCAGCT CATTTGCAAG TCAGAACCA ATACAGACCA ACTTGATTAT	660
	GATGTGGAG AAGAGCATCA GTCTCCAGGT GGCATTAGTA GTGAAGAGGA AGAGGAGGAG	720
30	GAAGAAGAGA TGTAATCAG TGAAGAGGAG ATACCATCA AAGATGATCC AAGAGATGAG	780
	ACCTACAAAC CCCACTTAGA AAGGGAAACC CCAAAGCCAC GGAGAAAATC AGGGAAGGTA	840
	AAAGAAGAGA AGGAGAAGAA GGAAATTAAA GTGGAAGTAG AGGTGGAGGT GAAAGAAGAG	900
35	GAGAATGAAA TTAGAGAGGA TGAGGAACCT CCAAGGAAGA GAGGAAGAAG ACGAAAAGAT	960
	GACAAAAGTC CACGTTTACC CAAAAGGAGA AAAAAGCCTC CAATCCAGTA TGTCCGTTGT	1020
40	GAGATGGAAG GATGTGGAAC TGTCCCTGCC CATCCTCGCT ATTTGCAGCA CCACATTAAA	1080
	TACCAGCATT TGCTGAAGAA GAAATATGTA TGTCCCCATC CCTCCTGTGG ACGACTCTTC	1140
	AGGCTTCAGA AGCAACTTCT GCGACATGCC AAACATCATA CAGATCAAAG GGATTATATC	1200
45	TGTGAATATT GTGCTCGGGC CTTCAAGAGT TCCACAAATC TGGCAGTGCA CCGGATGATT	1260
	CACACTGGCG AGAAGCATTA CAATGTGAGA TCTGTGGATT TACTTGTCGA CAAAAGGCAT	1320
50	CTCTTAATTG GCACATGAAG AACATGATG CAGACTCCTT CTACCAGTTT TCTTGCAATA	1380
	TCTGTGGCAA AAAATTTGAG AAGAAGGACA GCGTAGTGGC ACACAAGGCA AAAAGCCACC	1440
	CTGAGGTGCT GATGTCAGAA GCTCTGGCTG CCAATGCAGG CGCCCTCATC ACCAGCACAG	1500
55	ATATCTTGGG CACTAACCCA GAGTCCCTGA CGCAGCCTTC AGATGGTCAG GGTCTTCCTC	1560
	TTCTTCCTGA GCCCTTGGGA AACTCAACCT CTGGAGAGTG CCTACTGTTA GAAGCTGAAG	1620
60	GGATGTCAAA GTCATACTGC AGTGGGACGG AACGGGTGAG CCTGATGGCT GATGGGAAGA	1680

	TCTTTGTGGG AAGCGGCAGC AGTGGAGGCA CTGAAGGGCT GGTATGAAC TCAGATATAC	1740
	TCGGTGCTAC CACAGAGGTT CTGATTGAAG ATTCAGACTC TGCCGGACCT TAGTGGACAG	1800
5	GAAGACTTGG GGCATGGGAC AOCTCAGACT TTGTATTAA AAGTTAAAAA GGACAAAAA	1860
	AAAATCTAAA GCATTAAAA TCTAGTGAAG TAACTGAAGG GCCTGCTCTT TCCATTGTGG	1920
10	ATCACAGCAC ACACATACAT ACACCCTCCA CCTCCCCATC CCCTGTCTC CCTCTGTTC	1980
	TCCCCTTATA AAATTGATGT TGTCTTTACC AGAAAGGTAG AAAAAAGA AGCAGCAGCA	2040
	GCTCTTAAAG TGAGGGTTAT TCTCATACTC GGTCCAGCC ATCAGCAGAC TTCCTGCTCA	2100
15	TGGCAGATC CCCCTTTCCA ACCTGTAACCT CTGATGTGCT CTGGATCAGC TTTTAACTTT	2160
	TAATCATATA TTACTGTCTT CTAAATCCCT TCTCCTCCTC TACTGCTGCC CTATGGTTCT	2220
20	GGCTCCTACC CCCTGGGCA CACTTATCTT CAAATACCAT AGAATCTAA TCTCTGAAAT	2280
	CATAGCTCTC CAGTGGCTTT TAAAGAAAGC TGGTCTCAG CACTAACAA ATCACTACAA	2340
	TAGCCTAGTG CTTTTTTGGA AGCCTTTTTTA GGAAGAATG TTAGGTTTAT GGTAACTAGT	2400
25	ATGCTCTTTG AGATTTTTAC AGTGTGAAA CTTAAGAATT TTGAGAGGGT GAGGAGGGT	2460
	GTTCAGAATC TAAATTACAG ATAGATGATT GTTCTGTG AATTGTGTTT TTTTCTTTT	2520
30	TTTTGTGCCC TACCATTTC TTACATTTC CTGGGGCCC ATCTCTGGCT CCTGTCTTT	2580
	TGTTCTTTC TTTGCTTTAT CAGTTCATTC CAGCTCCCTG TTAGTGAAGG AACTGCTGT	2640
	TAGTGAAGGA ACAAAGTCTA TGAGTCTAA AATTTTAAGT CAAAGAAAC TGCTCTGTTT	2700
35	CCCCTTAGT AACACTTCTG AAGAGGAAA ACTTCAATAG CCAAAGTTAA TAATCCTATA	2760
	TAATAATTGC TTTGGCTTTC ACCTAAAATT CTGGGCATCA CAATTTCTTT GGGATAGAGG	2820
40	TTGTGTTGGG GAATAGATTG CTTATTGCTG TTCACTGGAG AGAAAAGGTA GTGTTTTTGT	2880
	ACAAGGTCAT ACCGCCAGAA GCCCCAAATC CTATTTTGGC TCATCTTCAG GTAAAGAGTA	2940
	ATTCCTATCC TGTGTGCCCTC AGAAGCTAGA ATCGAAGGCT TACCCTATTC ATTGTTTATT	3000
45	GTCAGAAATG CATGATGGCT CTGGAAAGA ATGACGTTT GCTGGAAAAA AAAAAAARA	3060
	CMGTTTGTGT TTCACAAACA TGGCTTATCA ATTTTTTCAA AGAATCTTT TTTCCCAAAA	3120
50	AGAGGAGTAA CAAATGTCA TTTCTGAAAG AGGCTTACTT TATACCAACT AGTGTACGCA	3180
	TTTGGGATGC CAGGAACAG AGAGTGAGAC ACCTACAATC ACCAGTCTCA AATGCGCTAT	3240
	TGTTTCTTTT CAGAGTGTG CAGATTGCC ATTTCTCCAT AATATGGGA TAGAAAATGG	3300
55	AATAAGATA GAAGGATGT AGAATATGCT TTCTGCCAA CATGGTTTGG AGTCGACTTT	3360
	GGTATATTGA CTAGATTGA AAATACAAGA TTGATTAGAT GAATCTACAA AAAAGTTGTC	3420
60	CTCCTCTCAG GTCCCTTTTA CACTTTTGA CTAAGTAGCA TCTATATTCC AACTTAGCT	3480

TTTTGTGCAC ACTTATCCTT TGTCTCCGTA AATTTTCATTT GCAGTGGTTA GTCATCAGAT 3540  
 ATTTTAGCCA CCTACACAAA AGCAAAGTGC ATTTTAAAAA ATCTTTCTGA GATGGGAGAA 3600  
 5 AATGTATTCT CCTTTCCTAT ACCGCTCTCC CAACAAAAAA ACAACTAGTT AGTTCTACTA 3660  
 ATTAGAAACT TGCTGTACTT TTTCTTTTCT TTTAGGGGTC AAGGACCTTC TTTATAGCTA 3720  
 10 CCATTTGCCT ACAATAAATT ATTGCAGCAG TTTGCAATAC TAAAATATTT TTTATAGACT 3780  
 TTATATTTTT CCTTTTGATA AAGGGATGCT GCATAGTAGA GTTGGTGTAA TTAAACTATC 3840  
 TCAGCCGTTT CCTGCTTTC CCTTCTGCTC CATATGCCTC ATTGTCCTTC CAGGGAGCTC 3900  
 15 TTTTAATCTT AAGTTCTAC ATTTTCATGCT CTTAGTCAA TTCTGTTACC TTTTAAATAA 3960  
 CTCTTCCCAC TGCATATTTT CATCTTGAAT TGGTGGTTCT AAATTCGTAA ACTGTAGTTG 4020  
 AGATACAGCT ATTTAATATT TCTGGGAGAT GTGCATCCCT CTCTTTTGTG GTTGCCCAAG 4080  
 20 GTTGTTTTGC GTAACGAGA CTCCTTGATA TGCTTCAGAG AATTTAGGCA AACACTGGCC 4140  
 ATGGCCGTGG GAGTACTGGG AGTAAATAA AAATATCGAG GTATAGACTA GCATCCACAT 4200  
 25 AGAGCACTTG AACCTCCTTT GTACCTGTTT GGGGAAAAAG TATAATGAGT GACTACCAA 4260  
 TCTAACTAAG ATTATTATAG TCTGGTTGTT TGAAATACCA TTTTCTCTC CTTTGTGTT 4320  
 TTTCCCACTT TCCAATGTAC TCAAGAAAAT TGAACAAATG TAATGGATCA ATTTAAATA 4380  
 30 TTTTATTCTT TAAAGCCTT TTTTGCCTGT TGTAATGTGC AGGACCCTTC TCCTTTCATG 4440  
 GGAGAGACAG GTAGTTACCT GAATATAGGT TGAAAAGGTT ATGTAAAAG AAATTATAAT 4500  
 35 AAAAGGGATA CTTTGCTTTT CAAATCTTGT TTTTCTCTTA TTCTAGGTAA GGCATATTAA 4560  
 AAATAAATAT GTAAAGAAGA AAAATAAAG TTGTCTTCAT GG 4602

40

(2) INFORMATION FOR SEQ ID NO: 75:

45 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1255 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

CGCGCCCCGG GCGGGCGGGT TTCTCTAACA AATAACAGA ACCCGCACTG CCCAGGCGAG 60  
 CGTTGCCACT TTCAAAGTGG TCCCCTGGGG GAGCTCAGCC TCATCCTGAT GATGCTGCCA 120  
 55 AGGCGCACTT TTTATTTTTA TTTTATTTT ATTTTTTTT TAGCATCCTT TTGGGGCTTC 180  
 ACTCTCAGAG CCAGTTTFTA AGGGACACCA GAGCCGAGC CTGCTCTGAT TCTATGGCTT 240  
 60 GGTGTGTACT ATAAGAGTAA TTGCCTAACT TGATTTTCA TCTCTTTAAC CAAACTGTG 300

	GCCAAAAGAT ATTTGACCGT TTCCAAAATT CAGATTCTGC CTCTGCGGAT AAATATTGTC	360
5	CACGAATGAG TAACTCCTGT CACCACTCTG AAGGTCCAGA CAGAAGGTTT TGACACATTC	420
	TTAGCACTGA ACTCCTCTGT GATCTAGGAT GATCTGTTC CCCTCTGGAT GAACATCCTC	480
	TGATGATCAA GGCTCCCAGC AGGCTACTTT GAAGGGAACA ATCAGATGCA AAAGCTCTTG	540
10	GGTGTATTAT TAAAACTA GTGTCACTTT CTGAGTACCC GCCGCTTCAC AGGCTGAGTC	600
	CAGGCCTGTG TGCTTTGTAG AGCCAGCTGC TTGCTCACAG CCACATTTC ATTTGCATCA	660
15	TTACTGCCTT CACCTGCATA GTCACTCTTT TGATGCTGGG GAACCAAAAT GGTGATGATA	720
	TATAGACTTT ATGTATAGCC ACAGTTCATC CCCAACCTA GTCTTCGAAA TGTTAATATT	780
	TGATAAATCT AGAAAATGCA TTCATACAAT TACAGAAATC AAATATTGCA AAAGGATGTG	840
20	TGTCCTTCTC CCCGAGCTCC CCTGTTCCCC TTCATTGAAA ACCACCACGG TGCCATCTCT	900
	TGTGTATGCA GGGCTATGCA CCTGCAGGCA CGTGTGTATG CACTCCCCGC TTGTGTTTAC	960
25	ACAAGCTGTG GGGTGTACG CATGCCTGCT TTTTCACTT AATAATACAG CTTGGAGAGA	1020
	TTTTGTATC ACATTATAAA TCCCACTGCG TCTTTTGTAT GGCCACATAA TAACTACTGC	1080
	ATAATATGGA TACGCCTTAT TTGATTTAAC TAGTTCCTA ATGATGGACT TTTAAGTTGT	1140
30	TTCTTTTTTT TTCTTTTTTT GCTACTGCAA ACGATGCTAT AATAAATGTC CTTATCAAAA	1200
	AAAAAAAAA AAAAAAAAAA AAAAAANCCC NGGGGGGGG CCCCAGGAAC NCAAT	1255

35

(2) INFORMATION FOR SEQ ID NO: 76:

40

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 475 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

	GGCAGAGAG AAATGTTTGA TTCTCTTTC TATTTTAAGG GATCTTCTCT CTGTGTGATG	60
50	TTGAAAACCTT ACCTTAGTGA AGATGTGTTT CAACATGCTG TTGTCCTTTA CCTGCATAAT	120
	CACAGCTATG CATCTATTCA AAGTGATGAT CTGTGGGATA GTTTTAATGA GGTCAAAAAC	180
	CAAACACTAG ATGTAAAGAG AATGATGAAA ACCTGGACCC TGCAGAAAGG ATTTCCCTTTA	240
55	GTGACTGTTC AAAAGAAAGG AAAGGAACTT TTTATACAAC AAGAGAGATT CTTTTTAAAT	300
	ATGAAGCCTG AAATTCAGCC TTCAGATACA AGGTACATGC CCTCTTCTTT TTCATGCCAT	360
60	CTCTTTTGCA CTCTCAGGTG GAAATATTTT GAAGTGTGTT ATAATCATAA GTTCTGTGTA	420



AACCTAACAA GATTATCCCT TCCTAAGAAT ACTTAACCTT CCTACCAAAT TAAAA

475

5

(2) INFORMATION FOR SEQ ID NO: 77:

(i) SEQUENCE CHARACTERISTICS:

10

(A) LENGTH: 465 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 77:

TTCTCTCTGC TCTTCGACTG CACCGCACTC GCGCGTGACC CTGACTCCCC CTAGTCAGCT	60
CAGCGGTGCT GCCATGGCGT GCGGCGGCG CGAACCRGCG TCGGGGCTCG CCGCGTGTG	120
GCTCTGGCGT TGCTCGCCCT GGCCTGTGTC GTGCCCGGG CCGGGGCGG GGCTCTCGAG	180
TGGTCTCGG CCGTGGTAAA CATCGAGTAC GTGGACCCGC AGACCAACCT GACGGTGTGG	240
AGCGTCTCGG AGAGTGGCGG CTTCGGCGAC AGCTCGCCA AGGAGGGCGC GCATGGCCTG	300
GTGGGCGTCC CGTGGGCGCC CGGCGGAGAM CTCGARGGCT KCGCGCCCGA CACGCGCTTC	360
TTCTGTCCCG AGCCCGGCGG CCGAGGGGCC GCGCCCTGGG TCGCCCTGGT GGTCTGTGGG	420
GCTGCACCTT TCAAGGACAA AGTGCTGGTG GCGGCGCNGA ANGAA	465

35

(2) INFORMATION FOR SEQ ID NO: 78:

(i) SEQUENCE CHARACTERISTICS:

40

(A) LENGTH: 1907 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

ACATGCAGCC CAACTACAGA TTCTTATGGA ATTCTCAAG GTTGCAAGAA GAAATAAGAG	60
AGAGCAACTG GAACAGATCC AGAAGGAGCT AAGTGTTTTG GAAGAGGATA TTAAGAGAGT	120
GGAAGAAATG AGTGGCTTAT ACTCTCCTGT CAGTGAGGAT AGCACAGTGC CTCAATTGTA	180
AGCTCCTTCT CCATCACACA GTAGTATTAT TGATTCCACA GAATACAGCC AACCTCCAGG	240
TTTCAGTGGC AGTCTCAGA CAAAGAAACA GCCTTGGTAT AATAGCACGT TAGCATCAAG	300
ACGAAACGA CTTACTGCTC ATTTTGAAGA CTGGAGCAG TGTACTTTT CTACAAGGAT	360
GTCTCGTATC TCAGATGACA GTCGAAGTGC AAGCCAGTTG GATGAATTC AGGAATGCTT	420
GTCCAAGTTT ACTCGATATA ATTCAGTACG ACCTTTAGCC ACATTGTCAT ATGCTAGTGA	480

60

	TCTCTATAAT GGTTCAGTA TAGTCTCTAG TATTGAATTT GACCGGGATT GTGACTATTT	540
	TGCGATTGCT GGAGTTACAA AGAAGATTAA AGTCTATGAA TATGACACTG TCATCCAGGA	600
5	TGCAGTGGAT ATTCATTACC CTGAGAATGA AATGACCTGC AATTCGAAAA TCAGCTGTAT	660
	CAGTTGGAGT AGTTACCATA AGAACCTGTT AGCTAGCAGT GATTATGAAG GCACTGTTAT	720
10	TTTATGGGAT GGATTCACAG GACAGAGGTC AAAGGTCTAT CAGGAGCATG AGAAGAGGTG	780
	TTGGAGTGT GACTTTAATT TGATGGATCC TAAACTCTTG GCTTCAGGTT CTGATGATGC	840
	AAAAGTGAAG CTGTGGTCTA CCAATCTAGA CAACTCAGTG GCAAGCATTG AGGCAAAGGC	900
15	TAATGTGTGC TGTGTTAAAT TCAGCCCCTC TTCCAGATAC CATTGGGCTT TCGGCTGTGC	960
	AGATCACTGT GTCCACTACT ATGATCTTCG TAACACTAAA CAGCCAATCA TGGTATTCAA	1020
20	AGGACACCGT AAAGCAGTCT CTTATGCAAA GTTGTGAGT GGTGAGGAAA TTGTCTCTGC	1080
	CTCAACAGAC AGTCAGCTAA AACTGTGGAA TGTAGGGAAA CCATACTGCC TACGTTCCTT	1140
	CAAGGGTCAT ATCAATGAAA AAAACTTTGT AGGCCTGGCT TCCAATGGAG ATTATATAGC	1200
25	TTGTGGAAGT GAAAATAACT CTCTCTACCT GTACTATAAA GGACTTTCTA AGACTTTGCT	1260
	AACTTTTAAG TTTGATACAG TCAAAAGTGT TCTCGACAAA GACCGAAAAG AAGATGATAC	1320
30	AAATGAATTT GTTAGTGCTG TGTGCTGGAG GGCACCTACCA GATGGGGAGT CCAATGTGCT	1380
	GATTGCTGCT AACAGTCAGG GTACAATTAA GGTGCTAGAA TTGGTATGAA GGGTTAACTC	1440
	AAGTCAAATT GTACTTGATC CTGCTGAAAT ACATCTGCAG CTGACAATGA GAGAAGAAAC	1500
35	AGAAAATGTC ATGTGATGTC TCTCCCCAAA GTCATCATGG GTTTTGGATT TGTTTTGAAT	1560
	ATTTTTTTCT TTTTTTCTTT TCCCTCCTTT ATGACCTTTG GGACATTGGG AATACCCAGC	1620
40	CAACTCTCCA CCATCAATGT AACTCCATGG ACATTGCTGC TCTTGGTGGT GTTATCTAAT	1680
	TTTTTGATA GGGAAACAAA TTCTTTTGAA TAAAAATAAA TAACAAAACA ATAAAAGTTT	1740
	ATTGAGCCAC AGTTGAGCTT GGAAAGTTTT TGTCAAATGC NGCAAGAGAT AACTCTTTTT	1800
45	ANGAAGTAGC ATATGTGAAC TATAATGTAA CAGTGAATAA TTTGTAAAGT TCGTATTTCC	1860
	CAACCTCTTT GGGAAATTACA CATATCAATA TAAACAAAAT ATAAAGT	1907

50

(2) INFORMATION FOR SEQ ID NO: 79:

55

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1168 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

GCTGGGGTGT CCCCKCSGCC ACCATCGTCA TCGCTTACTT GATGAAGCAC ACTCGGATGA 60  
 5 CCCATGACTG ATGCTTATAA ATTTGTCAAA GGCAAACGAC CAATTATCTC CCCAAACCTT 120  
 AACTTCATGG GGCAGTTGCT AGAGTTGAG GAAGACCTAA ACAACGGTGT GACACCGAGA 180  
 ATCCTTACAC CAAAGCTGAT GGGCGTGGAG ACGGTGTGT GACAATGGTC TGGATGGAAA 240  
 10 GGATTGCTGC TCTCCATTAG GAGACAATGA GGAAGGAGGA TGGATTCTGG TTTTITTTTCT 300  
 TTCTTTTTTT TTTTGTAGTT GGGAGTAAGT TTGTGAATGG AAACAAACTT GTTTAAACAC 360  
 TTTATTTTAA ACAAGTGTAA GAAGACTATA ACTTTTGATG CCATTGAGAT TCACCTCCCA 420  
 15 CAAACTGACA AATTAAGGAG GTTAAAGAAG TAATTTTTTT AAGCCAACAA TAAAAATATA 480  
 ATACAACITG TTTCTCCCCC TTTCCCTTTT AAGCTATTG TAGAGTTTAT GACTAAATAG 540  
 20 TCTGTGCAGG TTCATAGACC GAAGATACTA CACACTTTAA ACCAATTAAA AAGAACCAAA 600  
 AGTAAATAGA AAAGACATTG AATCACCAG GCCTGGGATC AACCTGGGCT GTCCACACAG 660  
 AAAACAAAAA CCCAACCAAA CCAAGCCCTG TTGTGCTCAC TGGTGCAAAG AGAAGATCAG 720  
 25 GGCAGCTTAA GTGGTCTAAG RATCCTTCAG GCATTCTTTA AGGAGAAAAA GGATACCTTT 780  
 GATTTTGTGT GTTTCATGCT CTGATTTTTT TTTTITTTTC CTCTCTGGG TTTAAGAGAT 840  
 30 TTTTITTTGAA ATAGTGAGGA ACTGACCATT ATATGCCTTC ACTGGCTTCT TGTGCAATAA 900  
 TATGATGTTT TAAGTGTGCA AACAAGTTAG AGCTGGCAGC TGAATGATAG ACAAATAGTG 960  
 CAAATTGCCC AGCTTGGAGA TAGAAAGGAA TTCAACAATA TATCAAATAC TTTCTTCCC 1020  
 35 ACCTTTTTC TTTTITTTTT TTTTITCTGA TTTGATCTG GTTACAGTGC CATAAACCTT 1080  
 GTTACATATG TATATCAGAA TGTAAGAAAA AAAAATTAT TTAATAATAT TTTTCGCAA 1140  
 40 AAAAAAANNA AAAAAGTCGA GGGGGGCC 1168

45 (2) INFORMATION FOR SEQ ID NO: 80:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1285 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

55 AGAAAATCAC ATCTTAACAA AGAAGTCTGT CTAAGACAGT ACATCTCCTG TTGAACTTGC 60  
 ATCTTTCCAC AGGACTTTCT GTTTTATAGG ATGAGACTAT TCTCTGCTTC ATCAAGGAAA 120  
 60 GAGAAATGTT CAGGGTTGTA GGGATGGCAC ACTTATTAGT TCTGCCTGTC TGAAAGGTTT 180

	CTGCAGGACA GTTTGGTCAG AGCTGCAATT CTTAGTCCAT GGTCTAATGC TTGAGTATCT	240
	CTTCTTTCCC TTCTCTGTCT CAGGAATCAG CTGAGAATTC ATTCGATTGT CATGCCTCTA	300
5	GCCCCTTACT GTGATTGTGT GGTGCACTT TCATTTGCTT TAGTTCTAGA ATCACCTGTT	360
	GACTCCTCAG ACTTCACCTA ACTTTGAAA CTCTCTTTTG GAGGCTTCTC ATTTCCTCCT	420
10	AATTCTGTGC TGCCTGAGCC CTAGAATTTT CCCACCAACG AATTATTCCA GGTAGATCCT	480
	AAGTTGCTGG ATCTAGTTGA TATTTAAACA ATATCTAGTT GATATTTCTC ATTCAGTTGG	540
	ATCCAGAAAC CAGTATCTCT NAAAAACAAC CTCTCATACC TTGTGGACCT AATTTTGTGT	600
15	GCGTGTGTGT GTGCGCGCAT ATGTATATAG ACAGGCACAT CTTTTTACT TTTGTAAAAG	660
	CTTATGCCTC TTGGTATCT ATATCTGTGA AAGTTTAAAT GATCTGCCAT AATGTCTTGG	720
20	GGACCTTTGT CTCTGTGTGA AATGGTACTA GAGAAAACAC CTATATTATG AGTCAATCTA	780
	GTTGGTTTFA TTCGACATGA AGGAAATTTC CAGATAACAA CACTAACAAA CTCTCCCTTG	840
	ACTAGGGGGA CAAAGAAAAG CAAACTGAC CATAAAAAAC AATTACCTGG TGAGAAGTTG	900
25	CATAAACAGA ATTAGGTAGT ATATTGAAGA CAGCATCATT AAACAGTTAT GTTGTCTCC	960
	TTGCAAAAAA CATGTACTGA CTCCCGTTG AGTAATGCCA AGTTGTTTTT TTTATTATAA	1020
30	AACTTGCCCT TCATTACATG TTTCAAAGTG GTGTGGTGGG CCAAATATT GAAATGATGG	1080
	AACTGACTGA TAAAGCTGTA CAAATAAGCA GTGTGCCTAA CAAGCAACAC AGTAATGTTG	1140
	ACATGCTTAA TTCACAAATG CTAATTCAT TATAAATTGT TTTGCTAAAA TACACTTTGA	1200
35	AACTATTTT CTGTATTCCA AGAGCTGAGA TCTTAGATT TATGTAGTAT TAAGTGAAAA	1260
	AATACGAAAA TAATAACAT TGAAG	1285

40

(2) INFORMATION FOR SEQ ID NO: 81:

- 45 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1290 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

- 50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

	TCTCCAGCCC CAATTTCTAC GCGCACCGBA AGACGGAGGT CCTCTTTCCT TGCCTAACGC	60
	AGCCATGGCT CGTGGTCCCA AGAAGCATCT GAAGCGGGTG GCAGCTCCAA AGCATTTGAT	120
55	GCTGGATAAA TTGACCGGTG TGTGTGCTCC TCGTCCATCC ACCGGTCCCC ACAAGTTGAG	180
	AGAGTGTCTC CCCCTCATCA TTTTCTGAG GAACAGACTT AAGTATGCCC TGACAGGAGA	240
60	TGAAGTAAAG AAGATTGCA TGCAGCGGTT CATTAAAATC GATGGCAAGG TCCGAACTGA	300

TATAACCTAC CCTGCTGGAT TCATGGATGT CATCAGCATT GACAAGACGG GAGAGAATTT 360  
 5 CCGTCTGATC TATGACACCA AGGGTCGCTT TGCTGTACAT CGTATTACAC CTGAGGAGGC 420  
 CAAGTACAAG TTGTGCAAAG TGAGAAAGAT CTTTGTGGGC ACAAAGGAA TCCCTCATCT 480  
 GGTGACTCAT GATGCCCGCA CCATCCGCTA CCCOGATCCC CTCATCAAGG TGAATGATAC 540  
 10 CATTGAGATT GATTTAGAGA CTGGCAAGAT TACTGATTTC ATCAAGTTCC ATTACCCAG 600  
 CCAGGTGGTC TCGTCACTC AGAGGCTCCG CAGACTCTG CCCAGGCCAG GACTGAGGCA 660  
 AGCCTCAAGG CACTTCTAGG ACCTGCCTCT TCTACCAAG ATGAACTCAC TGGTTTCTTG 720  
 15 GCAGCTACTG CTTTCTCTT GTGCCACCCA CTTTGGGGAG CCATTAGAAA AGGTGGCCTC 780  
 TGTGGGAAT TCTAGACCA CAGGCCAGCA GCTAGAATCC CTGGGCCTCC TGGCCCCSGG 840  
 20 GGAGCAGAGC CTGCCGTGCA CCGAGAGGAA GCCAGCTGCT ACTGCCAGGC TGAGCCGTCC 900  
 GGGGACCTCG CTGTCCCCGC CCCCCGAGAG CTCGGGAGC CCCAGCAGC CGGGCCTGTC 960  
 CGCCCCCAC AGCCGCCAGA TCCCCGCACC CCAGGGCGCG GTGCTGGTGC AGCGGAGAA 1020  
 25 GGACCTGCCG AACTACAAC TGAATCTCT CGGCCTGCGC TTGGCAAGC GGGAGGCGGC 1080  
 ACCAGGAAC CACGGCAGAA GCGCTGGCG GGGCTGAGGG CGCAGGTGCG GGGCAGTGAA 1140  
 30 CTTGAGACCC CAAAGGAGTC AGAGCATGCG GGGCGGGGC GGGGGCGGG GACGTAGGGC 1200  
 TAAGGGAGGG GCGCTGGAG CTTCCAACCC GAGGCAATAA AAGAAATGTT GCGTAACTCA 1260  
 35 AAAAAAAAAA AAAAAAANC TCGGGGGGGG 1290

## (2) INFORMATION FOR SEQ ID NO: 82:

40

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 684 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

45

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

50

TTTATTGTAT TCTGTAAC TAAGAACTCT ATTTWATTCT TTTTGGACT TGCTAAGTTG 60

TCTTTWATGG TTTTWAGTTC CATGCTGAAG TTTTCAGTAT TGACTTATCC CCTGAACAT 120

GAGTTGTTT ATAGACTCTR ATGATTCAAA AATCTTACAT CTTTGGTAG TCTCTTTCAT 180

55

TTGTYCACTG TTTCTGTTGA TTCTWACTCA TGGTATTTTA ATTCTTCGTT WTTTTTTTTTTC 240

TGTTWAGAWA CATTCTTTGA AAAATAATTT GGAGGAATAT TTGATTCTTA TGAACAAGGC 300

60

ATTACTCACC AGAGAAGATT TTTTGTGTYT ACCARGTGCC TARGAATGCT AACAGTCTGG 360

5 GAMCACATAG AMCACCAGGT GATGAGACAA TCCTGGGART CCTGTTTAC TTTGGSCCAT 420  
 CTTTCTCCC AACCTGTGG GAATARTCAT YCATATCCTA RCTGCAGGCT ARAAGGTGGT 480  
 10 TTATCAGAGC CCAACTCGA GGGCTCTGGG CTTTAGCTAC TGTACCCCA TCATAACTGA 540  
 GCTTCATGGA TTGATTCTCT TTTTATCTTT CAGATTTTCT TTTAAAAATC TTTGTTTTTT 600  
 TTTTCTTCC GAAAGATTCC CCAACATTA CCATTCCCA CCTTCCGTG AATTTTTTTG 660  
 15 GCTCTCATTT TGAATTTTTC AAGA 684

(2) INFORMATION FOR SEQ ID NO: 83:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2024 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 83:

25 CTGCAGGAAT TCGGCACAGC TGCCTGGAG GCTTCATCTT TGCCGCCGCT GCCGTGCGCT 60  
 TCCTGGGATT GGAGTCTCGA GCTTCTCTCG TTCGTTGTCG GCGGGGTCG CGCCCTTCTC 120  
 30 GCGCCTCGGG GCTGCGAGGC TGGGAAGGG GTTGGAGGG GCTGTTGATC GCCCGGTTTA 180  
 AGTTGCGCTC GGGGCGGCA TGTGGCCGG CGAGGTCGAG CGCCTAGTGT CGGAGCTGAG 240  
 CGGCGGACC GGAGGGGATG AGGAGGAAGA GTGGCTCTAT GCGATGAAA ATGAAGTTGA 300  
 35 AAGGCCAGAA GAAGAAAATG CCAGTGCTAA TCCTCCATCT GGAATTGAAG ATGAAACTGC 360  
 TGAAAATGGT GTACAAAAC CGAAAGTGAC TGAGACCGAA GATGATAGTG ATAGTGACAG 420  
 40 CGATGATGAT GAAGATGATG TTCATGTCAC TATAGGAGAC ATTAAAACGG GAGCACCACA 480  
 GTATGGGAGT TATGGTACAG CACCTGTAAA TCTTAACATC AAGACAGGGG GAAGAGTTTA 540  
 TGAACTACA GGGACAAAAG TCAAAGGAGT AGACCTTGAT GCACCTGGAA GCATTAATGG 600  
 45 AGTTCCACTC TTAGAGGTAG ATTGGATTG TTTGAAGAT AAACCATGGC GTAAACCTGG 660  
 TGCTGATCTT TCTGATTATT TTAATTATGG GTTTAATGAA GATACCTGGA AAGCTTACTG 720  
 50 TGAAAAACAA AAGAGGATAC GAATGGGACT TGAAGTTATA CCAGTAACCT CTACTACAAA 780  
 TAAAATTACG GTACAGCAGG GAAGAACTGG AAATCAGAG AAAGAACTG CCCTTCCATC 840  
 55 TACAAAAGCT GAGTTTACTT CTCTCCTTC TTTGTTCAAG ACTGGGCTTC CACCGAGCAG 900  
 GAGATTACCT GGGCAATTG ATGTTATCGG TCAGACTATA ACTATCAGCC GAGTAGAAGG 960  
 CAGGCGACGG GCAAATGAGA ACAGCAACAT ACAGGTCCTT TCTGAAAGAT CTGCTACTGA 1020  
 60 AGTAGACAAC AATTTTAGCA AACCACCTCC GTTTTCCCT CCAGGAGCTC CTCCCACTCA 1080

	CCTTCCACCT	CCTCCATTTC	TTCCACCTCC	TCCGACTGTC	AGCACTGCTC	CACCTCTGAT	1140
	TCCACCACCG	GGTTTTCTCTC	CTCCACCAGG	CGCTCCACCT	CCATCTCTTA	TACCAACRAAT	1200
5	AGAAAGTGA	CATTCTCTG	GTATGATAG	TCGTTCTGCA	CGTGCATTTT	CATATGGCAA	1260
	TGTTGCCTTT	CCCCATCTTC	CTGGTTCTGC	TCCTTCGTGG	CCTAGTCTTG	TGGACACCAG	1320
10	CAAGCAGTGG	GACTATTATG	CCAGAAGAGA	GAAAGACCGA	GATAGAGAGA	GAGACAGAGA	1380
	CAGAGAGCGA	GACCGTGATC	GGGACAGAGA	AAGAGAACGC	ACCAGAGAGA	GAGAGAGGGA	1440
	GCGTGATCAC	AGTCCTACAC	CAAGTGTTTT	CAACAGCGAT	GAAGAACGAT	ACAGATACAG	1500
15	GGAATATGCA	GAAAGAGGTT	ATGAGCGTCA	CAGAGCAAGT	CGAGAAAAAG	AAGAACGACA	1560
	TAGAGAAAGA	CGACACAGGG	AGAAAGAGGA	AACCAGACAT	AAGTCTTCTC	GAAGTAATAG	1620
20	TAGACGTCGC	CATGAAAGTG	AAGAAGGAGA	TAGTCACAGG	AGACACAAAC	ACAAAAAATC	1680
	TAAAAGAAGC	AAAGAAGGAA	AAGAAGCGG	CAGTGAGCCT	GCCCCGTAAC	AGGAGAGCAC	1740
	CGAAGCTACA	CCTGCAGAAT	AGGCATGGTT	TTGGCCTTTT	GTGTATATTA	GTACCAGAAG	1800
25	TAGATACTAT	AAATCTTGTT	ATTTTCTG	ATAATGTTTA	AGAAATTTAC	CTTAAATCTT	1860
	GTTCGTGTTG	TTAGTATGAA	AAGTTAACTT	TTTTTCCAAA	ATAAAAGAGT	GAATTTTTC	1920
30	TGTTAAGTTA	AAAATCTTTG	TCTGTACTA	TTTCAAAAAT	AAAAGACAG	CAATGACTTT	1980
	ATATCCAAAA	AAAAAAAAAA	AAAAAAAAAA	AAAAAAGGC	GGCC		2024

35

(2) INFORMATION FOR SEQ ID NO: 84:

- 40 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 931 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

- 45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

	CGCGCCMATA	GCCGGACGGG	GATCTGAGCT	GGCAGGATGA	ATGTGGGGGT	GGCACACAGC	60
	GAAGTAAACC	CCAACACCCG	AGTGATGAAT	AGCCGAGGCA	TCTGGCTGGC	CTACATCATC	120
50	TTGGTAGGAT	TGCTGCATAT	GGTCTACTC	AGCATCCCC	TCTTCAGCAT	TCCTGTTGTC	180
	TGGACCCGTA	CCAACGTCAT	CCATAACCTG	GCTACGTATG	TCTTCCTTCA	TACGGTGAAA	240
55	GGGACACCCT	TTGAGACTCC	TGACCAAGGA	AAGGCTCGGC	TACTGACACA	CTGGGAGCAA	300
	ATGGACTATG	GGCTCCAGTT	TACCTCTTCC	CGCAAGTTCC	TCAGCATCTC	TCCTATTGTG	360
60	CTCTATCTCC	TGGCCAGCTT	CTATACCAAG	TATGATGCTG	CGCACTTCCT	CATCAACACA	420

5 GCCTCATTGC TAAGTGTA CTGCGCGAAG TTGCCCCAGT TCCATGGGGT TCGTGTCTTT 480  
GGCATCAACA AATACTGAGG GATGGGTTTT GGGACAGCTC CATGGGCATG GGAAGGCAC 540  
10 TGAAACAGAG GACTATAAAA CATCCTTCTC TTATTCTCCA TACTGTCTTC TACACCTTTA 600  
AAGCCTGAGA ACTATACAAC CTTTCCCAGA CTCCAAGAA GAGAAGAGAT TGGCAAATGG 660  
GGCTCCTGGG CCCAGTCTG CTAGTGGCAA GTTCTTTGA ATCAGGAAGG CAGGTGAGGT 720  
15 AAGGCCAAA TCACTCTCTT CCATAGCAGG AAGCCATTG GGCAGCTCCT TTGGTGATTA 780  
CATCTTTCCA TATCTTTTAC ACTTACCACC TTCCAGCTCT GTTTTGCTGT GTATTTTCT 840  
TACAATAATT TTTTTCAGCT ATAGCTGCAG TTTAATCAGG ATGGGTAGAG AGCTGTCTC 900  
ATAAGGCTGG GGGTGGGAAG ATGAATACT G 931

20

(2) INFORMATION FOR SEQ ID NO: 85:

25 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 825 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

CGGGGCCGGC GGGGTCTTCA GGTACCGGG CTGGTTACAG CAGCTCTACC CCTCACGACG 60  
35 CAAACATGGC AGCGCAGAAG GACCAGCAGA AAGATGCCGA GCGGAAGGG CTGAGCGGCA 120  
CGACCTTGCT GCCGAAGCTG ATTCCCTCCG GTGCAGGCGG GGAGTGGCTG GAGCGGCGCC 180  
GCGCGACCAT CGGCCCTGG AGCACCTTCG TGGACCAGCA GCGCTTCTCA CGGCCCGCA 240  
40 ACCTGGGAGA GCTGTGCCAG CGCCTCTGAC GCAACGTGGA GTACTACCAG AGCAACTATG 300  
TGTTCTGTGT CCTGGGCCTC ATCCTGTACT GTGTGGTGAC GTCCCTATG TTGCTGGTGG 360  
CTCTGGCTGT CTTTTCGGC GCCTGTACA TTCTATCT GCGCACCTTG GATCCAAGC 420  
45 TTGTGCTCTT TGGCCGAGAG GTGAGCCAG CGCATCAGTA TGCTCTGGCT GGAGGCATCT 480  
CCTTCCCTT CTCTGGCTG GCTGGTGCGG GCTCGGCCGT CTCTGGGTG CTGGGAGCCA 540  
50 CCTGGTGGT CATCGGCTCC CACGCTGCCT TCCACCAGAT TGAGGCTGTG GACGGGAGG 600  
AGCTGCAGAT GGAACCGTG TGAGGTGTCT TCTGGGACCT GCCGGCTCC CGGGCCAGCT 660  
GCCCCACCCC TGCCCATGCC TGCTCTGCAC GGCTCTGCTG CTCGGGCCCA CAGCGCCGTC 720  
55 CCATCACAAG CCCGGGAGG GATCCCGCT TTGAAAATAA AGCTGTTATG GGTGTCATTC 780  
AGGAAAAAAA AAAAAAAGG GGGGCCCTC TAGGGGTCAA AGTTA 825

60



## (2) INFORMATION FOR SEQ ID NO: 86:

5

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1238 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

CATGTAAAAG GATGAAATGT GACTTCTGGT GTTTTTTTAT TTCTATGGAG GGACTTTCTG 60  
15 GGGACGGTTT CTGGCTCTCA GGCTCTGAGA AGCTGCAGTT TATGAGTGGC TCTGTGTGTG 120  
CTGCCACCTA CTGGAGAAGC CATAAGCTGC AGCTTTAGGA AAAGGGAACC CGGGGCAGAG 180  
TGTGGGGAAG TGGGATGGCA GCATGGCAGG GCTTTGGAAA ATGAGAGGTG AGAGTKTKTC 240  
20 CAGGAAGGTT GTAAGGAGAG GATGGATCCT GATACATGGA TTCAGGATCA TTAGGGTCCT 300  
GTCTGGGACA CTGGCCITCC TGCTTACCTG CTCTTTCCTT CCTCCTTGGT CGGAGGAGGG 360  
25 GCTGGCTCAC TGCTCTGGCT TCATTTTCCA GAGCTGCCTG CTGCAGTCAC ACTTAGGTCA 420  
TCTTCTCTCA CTTTTCTCCT TTTGCCGATT AGTGGACGTG ACAGAGATGT GAATGGGGCA 480  
GGGATGTCTT TGATGGCAT CAAGACTTTA GCTTCTGGTG CGCTGTGTCC CAGCTCTGAT 540  
30 TTCAGTTGCA GCCGTGATGG AMAGTTNGCA TGGAAGCTGA GACTCTCACT GACAGTGAAA 600  
CCCTCAAATG AACACAATCC CTGCTTTCCT GCCAAGGATC CTGTAGGGT NCCCCAGCT 660  
35 TCCCCACTTT TTTTCTGTGT CCTGACAAAG AAACACAGAG TAACTTGATT GCCCTGTGAC 720  
CTGGCCAGTT GCATTTCCTT TGCAGGCTTG AGCCCAAGCC AGAGCCTTGA AAAGGTATTC 780  
AGGTGTGTC CCAAAACACT GAAAAAACT GCCCTGGCCC TGAACCAAAT ACCTTGAACC 840  
40 CTCGTAAACT CCATACCCTG ACCCCCTTGT TTTGGATATA CCCAGGTAGA ACAACTCTCT 900  
CTCACTGTCT GTTGTGAGGA TACGCTGTAG CCCACTCATT AAGTACATTC TCCTAATAAA 960  
45 TGCTTTGGAC TGATACCCTT GCCAGTCTTT TGTCTTGGGC AATCTATACT TTTNCTCAGA 1020  
GGTTCCTAAG GCCTACTGAA GGGACTTAAC ATACTCTTAA TGGCTTTCCT CTCTCTTGT 1080  
TTACCTTATG CCCTCACTTC CTGAGTTAAC CTCCCAAATA CAGGATTCAC CTGTACCCAA 1140  
50 GCCCTTAGCT TCAAGAATAC AGGATCACCT GTACCCAAGC CCTTAGCTCA AGCTCTGCTT 1200  
TGGAAGAACC CAAACTAAGA CAGTGCTCCT GGTGCCCT 1238

55

## (2) INFORMATION FOR SEQ ID NO: 87:

60

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1460 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 87:

	ATTGCCTTCT GGTCCCTGGT GACACTGGGG TCATCCTTCA TCCCCGAGA GCATTCTTGG	60
10	CTGCTCCTCC TGACCCGGGG CCTGGTGGGG GTCGGGGAGG CCAGTTATTC CACCATCGCG	120
	CCCACTCTCA TTGCCGACCT CTTTGTGGCC GACCAGCGCG ACCGGATGCT CAGCATCTTC	180
15	TACTTTGCCA TTCCGGTGGG CAGTGGTCTG GGCTACATTG CAGGCTCCAA AGTGAAGGAT	240
	ATGGCTGGAG ACTGGCACTG GGCTCTGAGG GTGACACCGG GTCTAGGAGT GGTGGCCGTT	300
	CTGCTGCTGT TCCTGGTAGT GCGGGAGCCG CCAAGGGGAG CCGTGGAGCG CCACTCAGAT	360
20	TTGCCACCCC TGAACCCAC CTGCTGGTGG GCAGATCTGA GGGCTCTGGC AAGAAATCCT	420
	AGTTTCGTCC TGTCTTCCCT GGGCTTCACT GCTGTGGCCT TTGTACGGG CTCCCTGGCT	480
25	CTGTGGGCTC CGGCATTCTT GCTGCGTTCC CGCGTGGTCC TTGGGGAGAC CCCACCCTGC	540
	CTTCCCGGAG ACTCCTGCTC TTCCTCTGAC AGTCTCATCT TTGGACTCAT CACCTGCCTG	600
	ACCGGAGTCC TGGGTGTGGG CCTGGGTGTG GAGATCAGCC GCCGGCTCCG CCACTCCAAC	660
30	CCCCGGGCTG ATCCCTGGT CTGTGCCACT GGCCTCCTGG GCTCTGCACC CTTCCTCTTC	720
	CTGTCCCTTG CCTGCGCCCG TGTTAGCATC GTGGCCACTT ATATTTCAT CTTCATTGGA	780
35	GAGACCTTCC TGTCCATGAA CTGGGCCATC GTGGCCGACA TTCTGCTGTA CGTGGTGATC	840
	CCTACCCGAC GCTCCACCGC CGAGGCCTTC CAGATCGTGC TGTCCACCT GCTGGGTGAT	900
	GCTGGGAGCC CCTACCTCAT TGGCCTGATC TCTGACCGCC TGGCGCGAA CTGGCCCCC	960
40	TCCTTCTTGT CCGAGTTCCG GGCTCTGCAG TTCTCGCTCA TGCTCTGCCG GTTTGTGTGG	1020
	GCACTGGGCG GCGCACTTCC TGGGCACCGC CATCTTCATT GAGGCCGACC GCCGGCGGGC	1080
45	ACAGCTGCAC GTGCAGGGCC TGCTGCACGA AGCAGGGTCC ACAGACGACC GGATTGTGGT	1140
	GCCCCAGCGG GGCCGCTCCA CCGCGTGCC CGTGGCCAGT GTGCTCATCT GGAGAGGCTG	1200
	CCGCTCACCT ACCTGCACAT CTGCCACAGC TGGCCCTGGG CCCACCCAC GAAGGGCTG	1260
50	GGCCTAAACC CCTTGGCCTG GCCAGCTTC CAGAGGGACC CTGGGCCGTG TGCCAGCTCC	1320
	CAGACACTAC ATGGGTAGCT CAGGGGAGGA GGTGGGGGTC CAGGAGGGG ATCCCTCTCC	1380
55	AACAGGGGCA GCCCAAGGG CTCGGTGCTA TTTGTAACGG GATTAAAATT TGTAGCCAGA	1440
	AAAAAAAAA AAAAAAAAAA	1460

60

## (2) INFORMATION FOR SEQ ID NO: 88:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1395 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 88:

5 CAGGTGCAAA GTGGGAAGTG TGAGTCCTCA GTCTTGGGCT ATTGGGCCAC GTGCCTGCCG 60  
GACATGGGAC GCTGGAGGGT CAGCAGCGTG GAGTCTGGC CTTTTCGTC CACGGGTGGG 120  
15 AAATGGCCA TGCCACGGC GGAAC TGGG ACTCAGGCTG CCCCCGGCC GTTCTCATC 180  
CGTCACCGG AYTCTGGGC GCTCGCACTG GCGCTGATGT AGTTCTCTGA CCTCTGACCC 240  
20 GTATTGTCTC CAGATTAAAG GTACGACATT TGGAGGCCCC AGCGAGAAAC GTCACCGGGA 300  
GAAACGTCAC CGGGCGAGAG CGKCCCGCT GTGTGCTCCC CCGGAAGGAC AGCCAGCTTG 360  
TAGGGGGGAG TGCCACCTGA AAAAAAATT TCCAGGTCCC CAAAGGTGA CCGTCTTCCG 420  
25 GAGACAGCGG ATCGACTACC ATGTGGGTGC CCACAAAAT TYCACCTYTG AGTCCTCAAC 480  
TGCTGACCCC GGGTTCAGTT CCAGAGAGAA GGACTCCCTC CTGCTTGGA GAGACCTCAC 540  
ACCGTCATCA CGATGCCAAC GGCTCTGAAG GTGGATGGCA TTCCTGGTG GATTCATCAC 600  
30 TCCCGCATCA AAAAGGCCAA CRGAGCCCA CTAGAAACAT GGGTCCCCAG GGCTGGGTCA 660  
GGCCCCTTAA AACTGCACCT AAGTTGGGTG AAGCCATTAG ATTAATCTT TTTCTTAATT 720  
35 TTGTAAACA ATGCATAGCT TCTGTCAACT TATGTATCTT AAGACTCAAT ATAACCCCT 780  
TGTTATAACT GAGGAATCA ATGATTGAT TCCCCAAAA CACAAGTGG GAATGTAGTG 840  
TCCAACCTGG TTTTACTAA CCTGTCTT AGACTYTCCC TTCTCTTAA TCACTCAGCC 900  
40 TTGTTTCCAC CTGAATTGAC TCTCCCTTAG CTAAGAGCG CAGATGGACT CCATCTTGGC 960  
TCTTTCTNACT GGCAGCCGCT TCTYCAAGG ACTTAACTTG TGCAAGCTGA CTCCAGCAC 1020  
45 ATCCAAGAAT GCAATTAAT GATAAGATAC TGTGGCAAGC TATATCCGCA GTTCCCAGGA 1080  
ATTCTGTC AA TTGATTACAC CMAAAGCCC CGCGTCTATC ACCTTGTAAT AATCTTAAAG 1140  
COCTGCACC TGGAATATT AACGTCTCTG TAACCATTTA TCCTTTTAA TTTTTCCTT 1200  
50 ACTTTATTTT GTAAAATTG TTTTAACTAG ACCCCCCCTC TCCTTTCTAA ACCAAAGTAT 1260  
AAAAGCAAAT CTAGCCCTT CTTAGGCCG AGAGAATTTC GAGCGTAGC CGTCTCTTGG 1320  
55 CCACCAGCTA AATAACGGA TTCTTCATGT GTAAAAAAA AAAAAAAAAA CTCGGAGGGG 1380  
GGCCCGGTA CCAA 1395

## (2) INFORMATION FOR SEQ ID NO: 89:

## (i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 1186 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 89:

GGCACGAGCC GGCAAGCCGA GCTAGGGTGA AAAC TGGGGG CGCACCAGGA TGTNNGACAG 60  
AAAAGCAGAA GATGAGACTC TGTTCATTCA CTTTTCCTAG GCCCATCTG TGGTCATCTT 120  
TCCCCCTCCC ATCATACCTC CTCCTCCTG GAGCCTCTGC CGGCTGGCT GTAATGGTGG 180  
CACTTACCTG GATATTTTCA TGGGAGGATG AAAGGCGAGA CTCACCCTAC GCGGTGGGAC 240  
20 AGATGGGGAG AGGAAAAAGG CAGAGATGGC CAGGAGAGGG GTGCAGGACA AACCAGAGAG 300  
GTGGGTTCAG GGGAAAAGG TGGGAGAGAA GAGGGGTGCA GGCCCTGCAG GCCGGTTAGC 360  
CAGCAGCTGC GGCCTCCCGG GGCCCTTGGC ATCCAACCTC GCAGACAGGG TACCAGCCTC 420  
25 CTGGTGTGTA TCATAGGATT TGTTCACATA GTGTTATGCA TGATCTTCGT AAGGTTAAGA 480  
AGCCGTGGTG GTGCACCATG ACATCCAACC CGTATATATA AAGATAAATA TATATATATA 540  
30 TGTATGTAAA TTATGGCAGC AGAAATTATA GCACTGAGGG CCCTGCTGCC CTGCTGGACC 600  
AAGCAAACT AAGCCTTTTG GTTTGGGTAT TATGTTTCGT TTTGTTATTT GTTTGTTTTT 660  
GTGGCTGTG TTAATGCTG ATAGCACAAG TGCCAGTCGG ATTGCTCTGT ATTACAGAAT 720  
35 AGTGTMTTTA ATTCATCAAT GTTCTAGTTA ATGCTCTACCT CAGCACCTCC TCTTAGCCTA 780  
ATTTTAGGAG GTTGCCTAAT TTTGTTTCTT CAATTTTACT GGTTACTTTT TTGTACAAAT 840  
40 CAATCTCTTT CTCTCTTTCT CTCCTCCCA CCTCTCACC TTGCCCTCTC CATCTCCCTC 900  
TCCCGCCCTC CCTCCTCCC TCTGGCTCCC CGTCTCATTT CTGTCCACTC CATTCTCTCT 960  
CCTCTCTCC TGCTCTCTG TGCCCCCTCC CCAGCCCACT TCCCGAGTT GTGCTTGCCG 1020  
45 CTCTTATCT GTTCTAGTTC CGAAGCAGTT TCACTCGAAG TTGTGCAGTC CTGGTTGCAG 1080  
CTTTCCGCAT CTGCTTCGT TTCGTGTAGA TTGACGGTT TCTTTGTAAT TTCAGTGTTT 1140  
50 CTGACAAGAT TTAATAAAAA AAAAAGGAAA AAAAAAAAA AAAAAA 1186

## 55 (2) INFORMATION FOR SEQ ID NO: 90:

## (i) SEQUENCE CHARACTERISTICS:

- 60 (A) LENGTH: 1821 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 90:

5	AAAACATGCT TTCAGGGCGT CCCCTATGTA TTCGGGGGGC CCACGGACAC TCAGGCTGGA	60
	KATCCGTCCT CACTGCGCTC AAGATGGCCT CAGCAGACAC CAGTTACCCA GCTGAAAGTC	120
	ACAATCCCTC CCAGAAGTCT CCCAACACTA GTGCTGACCA GAGGTGGGGC TCTCAGGCTA	180
10	GGAGTTTCAC ACACAATGAC AGGCTGCTGG GGGACATGTC AGGACCCCTT TTCCTYTCCT	240
	CTCCATGCTA GAAGCCAGCC CTAGGMAGCT GCAGTTACTC CCTGTGACTC AGCAGCAGGC	300
15	TGATTCAACA CAGCTGCCCA CACAAAGCCA GTGGTAATAC ATCTGTTTAC CTPTCCCTAT	360
	CACCCAGACA CAAGCCCCTT TCCCAGGTCA AACCACAGGC OGATGCATCT CCAGTTTGAC	420
	AGTCAAATCA CTACTTCCAT TGCTACTTTA GATCAGCCAA AGTGGTGACT GCTGCAGTGT	480
20	GTGGCTATCC CTACAAGGCC CACCCAAGGG ATGCCCAAAG CCCAACCTTC TCCAGGGCTG	540
	CAGCAGNAGC AACCCACCA GCCTAAGTCC AGCAGAGGAC CTCCCACCCA ATGTCTTGTT	600
25	CTAATTAGAA GGGGAAGTTA GCCACAGAAA ATCAACTTAT CTATAATTAC AAAATTCCTCT	660
	TGACTCACCT TAAAGTTCCT ATTGACATCT ACTGCTTTTA AACCTATTTC AAAACTCTGA	720
	TACTAAACA AATGACACTC TAAGAAAGTT TGGGAGCCCC ATGCTGAGAA CCATTTCTGT	780
30	GCAGTGAGGA TGTTCACAGA AGCTACTTAC CTACATGTGA ATGTGCCATT TTCTTTCCTT	840
	TTGTAGAGAA AATCCCCTTT ACTTTTGGGA ACAGTAATGG CAGCTTCTAG TACAGCCATT	900
35	ACAGTTTCAT ATGAGAAAAA TTAAGAATAA CTATAAAATT GTTAAATAT CCAATAATGG	960
	ATAATGATGG CCAGAAGATT TAACATACAA AGTAATTCTC AATGTAAAGC TATTCACTC	1020
	TTCCAGGTG AATGCCCTGT AACCCACCT GACCTCCAC ATCATCTTCA AAAAGCAGTT	1080
40	TCTCTGTTCC CCATGATTCT CCTATAAGGT AACTCTTTAG TCCTCCATT AGCACATTTT	1140
	AAATCCTCCA AAGAATAAGT ATCATGTGAT TATTTTAGCT TTACAAAAAA AAAGTTGAAT	1200
45	GGCGTTTAT TTTCATGGCC TATAAGCAGG TACCTTAGTA GGCAGATAT AGGAAAAACA	1260
	AATTAGAGCA AAACAAATCC TCTACAAATC CAAGGCAGGA AAAGTGGTGG CAGAGTGAAT	1320
	CATTCTCTG TCCCTCCCAT CAGGTCAAAT CAGGAGGCTG CAGTGAATGC CTGTTCTTTG	1380
50	AATGTGTAGC AGTGTTCCT GTAACCTTT AAAACTTGGC TATAGGCTGT TTAGCACAGT	1440
	ACAGATTAAA GATACAGTTA CGTAAACAGC AAAGTAATTT TATAGTGCTT CATCCATTTA	1500
55	TCATGCTTTG GTTTGCTAAT TTTTCACAT ACCTTTTCT ATCACAGTCT GTTGCTTTTG	1560
	TACACATTTT TCATATTGGG GTTCGACAGG TAAACACAAA CTGCTATTTC AGTAGAAAAA	1620
60	GTTATTGTTA TGAATATTA AACCCAATAA ATTGTATAAA GGGTAAAAAA AAAAAAAAAA	1680

AAAAAAAAA AAAAAAAAAA AAAAAAATTC CTGCGGGCCG CANGCTTTTT CCCTTTGGGT 1740  
 GAGGGGTAT TTTNGGCTG GGCAGTGGC CCTTCGTTTT TACAACGTCG TGANGGGGGG 1800  
 5 AACCCGGGGG GGGTTTCCCC C 1821

10 (2) INFORMATION FOR SEQ ID NO: 91:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 862 base pairs  
 (B) TYPE: nucleic acid  
 15 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 91:

20 TGCCCTTTTT CCCACCGATT CGGGGCGTGG TGAAGGTGGG AGATGTGAAC TCCAATTAAG 60  
 GGACTGGAGA GAGGTGAAGA ATTTTGCAGG TGGGAGATTT GGATTTGAAT GTGGACTTGT 120  
 AAATGACTTG ACCTTGCCAT CTGTGTTCAA GGTACGGTT TGCTGTGGG TTCTGCGGAG 180  
 25 AGCTTACTCA CCCCGGAGTC TTTTCTTTCT CTGTCTCAA GAAGAGCCCT GTTGGTGCTT 240  
 TACCACCGCT TGGAGTCTCC CGAGGACACA AACAGGCAGA GAGGGAAGTG TAGGGAGAGT 300  
 30 TCTTTCCTGT TTTCTGTGCT TTCTTTTTA CAGGACTCCC GGAAGGCCAC TCATGGCCAT 360  
 GCCAGGAGCT TTCTCAGAAA CAGTCATAAA CGATCTCTTG AGTCTCTTTC TTGTCTCTCC 420  
 AGCTGAGCTT TCTTATTTCA CCCTTTCTGG TGTCTATAGG AATGCATGAG AAGACCCTGG 480  
 35 GACGTTTTTC TGCTCTCTTC TGGCCCTCCA TGGAGCCATG GGCTCGGCC TCGGCGGCTC 540  
 CTCACCTCA CAATTTATTT CCTCCTCCCG TGCCAGCCCT TCTTTTGTGT CTGAAACCGG 600  
 40 TTTTAAATG TGAATCTCCC AGAGAAGAAG CCGCTGGCTG TATGAAACTT GACGGCGCTT 660  
 TTGTAAGGTG CCACCCCAA ACTTTAAGGT AGCTAAACCA ATTTTAAAA GATTCATGG 720  
 CTGTTCATC CTCCAGATGT AGCTATGAT GTACACTTCG CAACGGAGTG TCTGAAATG 780  
 45 TGGTGGTCCT GATTTATAGG ATTCATAAT TAAATGTCT GCTGAATAA AAAAAAAAAA 840  
 AAAAATCGA GGGGGCCCG GT 862

50

(2) INFORMATION FOR SEQ ID NO: 92:

- 55 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 696 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 60 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 92:

	CTGAGGCGAG TGAAGTGGAC TCTGAGGGCT ACCGCTACCG CCACTGCTGC GGCAGGGGCG	60
5	TGGAGGGCAG AGGGCCGCGG AGGCCGAGT TGCAAACATG GCTCAGAGCA GAGACGGCGG	120
	AAACCCGTTT GCCGAGCCCA GCGAGCTTGA CAACCCCTTT CAGGACCCAG CTGTGATCCA	180
10	GCACCGACCC AGCCGGCAGT ATGCCACGCT TGACGTCTAC AACCCCTTTG AGACCCGGGA	240
	GCCACCACCA GCCTATGAGC CTCCAGCCCC TGCCCCATTG CCTCCACCCT CAGCTCCCTC	300
	CTTGCAGCCC TCGAGAAAGC TCAGCCCCAC AGAACCTAAG AACTATGGCT CATAAGCAC	360
15	TCAGGCCTCA GCTGCAGCAG CCACAGCTGA GCTGCTGAAG AACAGGAGG AGCTCAACCG	420
	GAAGGCAGAG GAGTTGGACC GAAGGAGCGA GAGCTGCAGC ATGCTGCCCT GGGRGGCACA	480
20	GCTACTCGAC AGAACAATTG GCCCCCTCTA CCTTCTTTTT GTCCAGTTCA GCCCTGCTTT	540
	TTCCAGGACA TCTCCATGGA GATCCCCCAA GAATTTGAGA AGACTGTATC CACCATGTAC	600
	TACCTCTGGA TGTGCAGCAC GSTGGNTCTT CTCCTGAAYT TCMTGSGCTG CCTGGCCAGT	660
25	TCTGTGTGGA AACCAACAAT GCGGAGGCTT TGGGTT	696

## (2) INFORMATION FOR SEQ ID NO: 93:

## (i) SEQUENCE CHARACTERISTICS:

- |    |                             |
|----|-----------------------------|
|    | (A) LENGTH: 1886 base pairs |
| 35 | (B) TYPE: nucleic acid      |
|    | (C) STRANDEDNESS: double    |
|    | (D) TOPOLOGY: linear        |

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 93:

40	CAGGCCACTG ACGCTTCTTT GCGAGGGATG CAGGAGGTCC TACAGAGAAA GCGGCTTCTT	60
	GCATKTCAGA GGGCCACAG CCTGTACCCC ACAGATCACC AAGCAGCTTT CTACCTGGCT	120
45	CTGCAGCTTG CCATCTCCAG ACAGATCCCA GAGGCTCTGG GGTATGTCCG CCAAGCTCTT	180
	CAGCTTCAAG GTGACGATGC CAACTCCCTG CACCTCCTTG CCTCCTGCT GTCAGCACAG	240
	AAGCATTACC ATGACGCTCT GAACATCATC GACATGCCCC TGAGTGAATA CCCAGAAAAT	300
50	TTCACTACTAC TGTTTTCCAA AGTGAAGTTG CAGTCACTCT GCGAGGCCC GGACGARGCA	360
	CTGCTGACTT GTAAGCACAT GCTGCAGATA TGGAAATCCT GCTACAACCT CACCAACCCC	420
55	AGTGATTCTG GACGTGGGAG CAGCCTCTTA GATAGAACCA TTGCTGACAG ACGACAGCTT	480
	AATACAATTA CTTTGCCAGA CTTGAGGAT CCGGAGACAG GCTCCGTCCA TGCCACATCG	540
	GTAGCAGCCT CAAGAGTGA GCAGGCACTG TCGGAAGTGG CTTCGTCTCT GCAGAGCATG	600
60	CCCCTAAGCA GGGCCCCTG CACCCCTGGA TGACGCTGGC ACAGATCTGG CTCCATGCAG	660

	CTGAAGTCTA TATCGGCATC GGAAGCCTG CAGAAGCCAC AGCCTGTACC CAAGAAGCTG	720
5	CCAACCTCTT CCCAATGTCC CACAATGTCC TCTACATGCG CGGCCAGATT GCTGAGCTCC	780
	GGGGAAGCAT GGACGAGGCG CGGCGGTGGT ATGAAGAGGC CTTAGCCANT CAGCCCCACC	840
	CACGTGAAGA GCATGCAGCG ACTTGGCCCT GATCCTTCAC CAGYTAGGCC GTTACAGTYT	900
10	GGCGGAGAAG ATCCTCCGGG ACGCGGTGCA GGTGAACCTG ACAGCCCACG AGGTCTGGAA	960
	CGGGCTGGGC GAGGTCTCC AAGCTCAGGG CAACGATGCG GCGGCTACG AGTGCTTCCT	1020
15	GACAGCCTTG GAGCTGGAGG CCAGCAGCCC CGCGTGCCC TTCACCATCA TCCCCCGGT	1080
	GCTCTGAGCA GGCCTCTGCC AGCCTCACCT GCGCTCAGC CTNCAGAGGC CTTGCCGGG	1140
	ACCAGGGCTT GTGCCATGCG CCCAAGGGGA TGAATCTGCC GCACTGAGGC CAGGGACGAG	1200
20	TGTTCACTGG GCCACAGTGA ACCAACCAAA CCAACCCGA ATCATCGCTC TCGCCATGTG	1260
	CGTTTCTCTT GTTTTTTTT CCAGCCCAAT GGTAGTTCT GAACCTATTG ACATTGTTCA	1320
25	AAATGGATCA TGTGCCATAT TTGTAGT GACATCTGAG TTTTCAGTAA AATGATTATG	1380
	GAATTAATCA GCAAATGTAG AAGAATATAT TCAAAGTTAA AATTCAGTGG CAGCACAGAT	1440
	TATTTTATC AGAGCTGTAA AGAAAACAAC TGTCTTTTC TCCCCACCAC CCTCTCTGCC	1500
30	CCACTTTGGC CCAGAAACCA AATGTGAAC TCTGTCTCC CACCTCAGCA CTAGTCCATG	1560
	CCAGGACACC AGCTGACAAT TTCTTGTTT TACTGTCAAT AATGTACCA TGTGATCAAT	1620
35	TACTGTCTC ACTTAGAACA AAGCTGAGT CCGAGAATAT TTATATTTTA CCAATATATG	1680
	CCTGTTACAA GAGAAGGAAA TATGAGTTAT TTAAGTTTAA CTTTTTTATG TGAATTCAGA	1740
	GTTTATTTAT CGAGGGAAAT ATGTACAAAG AAGCTTCAA TGAATATTT ACCGACATTC	1800
40	CTTATACATG ACAGACACTT GGCTACATGG GAAGATGATG TTAATAATAA AATGATTTTT	1860
	AAATGGAAAA AAAAAAAAAA AAAAAN	1886

45

(2) INFORMATION FOR SEQ ID NO: 94:

50

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1774 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 94:

60

CTCAGCTACC GTATACAGTA GGACATAACC CCATTTTACA TGCACTACAC TGAGACTTGC	60
CTCCTCTCCC CCCACATTGA AGATGTTCTT TTTTCATAAC TATATACTAT TCCATTGCAT	120



	GAATATCTG TAATTTATTT AATCCCCTAT GGATTGATAA TTAGGTTTCAT TATAGATAGA	180
	AGTGTAATTA ACATTCTCTG ACATGTATTT TGCTACTTGT GTGGGTATTT CTGTAGGATG	240
5	AATACTAGA AATTTATTTG ATCAGGTTTC ACATTTGCAG TTTTGAAAAC TACTACCAAA	300
	AAGATTTTAC CAATTTACAA CTCCATCATT AGTAAGAATG CCTGTTTGCC TATAGTCTGC	360
10	CAACCTGAA TCCTTAAAAA TTTTGGCCAA TCTGGTAGGC AAAATTTCTT TCTTTTCTTT	420
	GAATATTAAT GAGGAGGAAC ATCTTTTCAT GTTCTTTGGC CATTTGCATT TCCTATTATG	480
	AATTGCTTTT GCCCATTTTC CTTTTTTTAA TTATGAAAGT CTAATGACTA CCTTCTCATT	540
15	GTATAAAAAA CACAGTTCTT TGAATAGAGA GACCCTTTTC TCCAATGCTA CCAATCACAT	600
	TCCACTTACC ACAGTTTAAC ATACATCCTC TAGTCACCTT TCCGTACGAA TATACATACA	660
20	CATAAAAACA CTTTTTACAT AAATAGGATC TCATATCTG TAGCTTTTAA AAATTTGGT	720
	CTCAAAAAA GATAACAGGT CTTTAAATTT CTTTAATGGT TGAATATGAT TAAATACTAT	780
	GAAAATGCCA TTATTTATTTC CCTTAATTTT TTTCCCTCTG CTATTACATT GCCAAAGTAA	840
25	ACATCCTATT CAGATGTCTT TGTGCATGTG TGTGAATATT TCTTTAGTCT GGAGTCCAGT	900
	AAGGTGGATT TTTGGATCAA AGGGTTTGT CTCTGTCCAC CTTCACTCTT CCCAAAGGCC	960
30	TTCATAACTG TATTTTCACC AAGTGTATGG AGAATGTTCA TTTCCCCATA TAACCATACC	1020
	TACACTTGAT AGTTTTTATC TGTGGGGCGA AAAAGAACCT TTCTTTATTT TGCATTTCCC	1080
	TGATTATAAA AAAAAATGGT GAGATTGGGG TTATTTTCAT GTTATTTGGC CATTTATAGT	1140
35	TTACTGTGGA TTGTTTGTAT CCTTACCTG CTTTCTATTG GGTTATGTGT GGATATATTG	1200
	TTTTTATTTG TTCAGCATCT CCTTCCCAT CTTCTGGTAA CACAACCTTT ATTTATTTGT	1260
40	GGGGAACCTA TTCCCTGTGG CTTAGGTGAG CATGTGACCA GGCCTGGCCT CTGAGTCCC	1320
	ACAGCTTCCT AGCCACAGTG ATAAAAGAAT GGGTATATAA CTTAAGCCAG GCTAAGGAAA	1380
	GCCCTTAACA GAATTTCTGC TGGAACTACT GGAAAGAAGG CTTTATGGAG ATCCCAGGAA	1440
45	CCAAGGACCA TGTAAGCCTG AATTTGTGCC ATGTGGAGAG AGTCTGTCTG AGGAGAACT	1500
	CGGATGCTAG CAGAAATGGA AAGAGAACTA AGTTCTGATG TCATTTTCTT GGAGGCCCTA	1560
50	GATCCAGCTG TGCCTAAAGC CTGCCCTACT CCGGACTTTA AAGTTTGTG AGCCAATAAA	1620
	GTCCCTTTCT TGTTTAAGAT AATGAATTG AGTTTCTGTT CTGATTAATA TAGGTTATTT	1680
	GTATTTTCTT ATTGATTTGT AGAAAACCTT TGTAATTTTA AATTTCTAGAC TTTATGCACT	1740
55	ATATAAGTTA ATAAAATTAG CATGGCCCTC CATG	1774

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2503 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 95:

10	GGCAGGAGCG AAGGCAAGGG GGCACCAGCT CAGGACTGCA TCTGCCTGCC ATTTCCCTTC	60
	CACTCCTCCT TTCTGGAGTC TGACATTAGA AAGCCAGCGA GAAGGAAGAT TCAAACAACC	120
15	AACCTGATT TCCTGCTTCT CCTTTTCATG AGTGTTCCTG TGGTCTCTGC ACCTCCTTTC	180
	TGTCCCCCGG CAGAGGGCAG TAGAGATGGC CGGCCCAAGG CCTCRGTGGC GCGACCAGCT	240
	GCTGTTTCATG AGCATCATAG TCCTCGTGAT TGTGGTCATC TGCCTGATGT TATACGCTCT	300
20	TCTCTGGGAG GCTGGCAACC TCACTGACCT GCCCAACCTG AGAATCGGCT TCTATAACTT	360
	CTGCCTGTGG AATGAGGACA CCAGCACCTT ACAGTGTAC CAGTTCCTG AGCTGGAAGC	420
25	CCTGGGGGTG CCTCGGGTTG GCCTGGGCCT GGCCAGGCTT GGCGTGACG GGTCCCTGGT	480
	CCTCACCTTC TTTGCCCCCC AGCCTCTCCT CCTAGCCCAG TGCAACAKTG ATGAGAGAGC	540
	GTGGCGSCTG GCAGTGGGCT TCCTGGCTGT KTCCTCTGTG CTGCTGGCAG GCGGCCTGGG	600
30	CCTCTTCTTC TCCTATGTGT GGAATGGGTC ARGCTCTCCC TCCGGGGGCC TGGGTTTCTA	660
	GCTCTGGGCA GCGCCAGSC CTTACTCATC CTCCTTGCTTA TAGCCATGGC TGTGTTCCCT	720
35	CTGAGGGCTG AGAGGGCTGA GAGCAAGCTT GAGAGCTGCT AAAGGCTTAC GTGATTGCAA	780
	GGGTTTCAGTT CCAACCATGG TCAGAGGTGG CACATCTGCT CAGCCATCTC ATTTTACAGC	840
	TAACGCTGAT CTCCAGCTCC AGCGATGGAA CCCACTACAG AGGAGGTGGG GCCCCTGTGT	900
40	CAAAGAGGCC GAGGGGCAGC AAGGGCAGMC AGGGCACCTG TGACTTCTTA GTACAAGATT	960
	GTCTGTCTTT CAGGACTTCC AAGGCTCCCA AAGACTCCCT AAACCATGCA GCTCATTGTC	1020
45	ACACCAATTC CTGCTTTAAT TAATGGATCT GAGCAAATCT TCCTCTAGCT TCAGGAGGGT	1080
	GGGAGGGGAG TGATTGCTGT CATGGGGCCA GACTTCCAGG CTGATTGGCC AAATGCCAAA	1140
	ATGAAACCTA GCAAAGAACT TACGGCAACA AACGAGGACA TTAAAAGAGC GAGCACCTCA	1200
50	GTGTCTCTGG GGACATGGTT AAGGAGCTTC CACTCAGCCC ACCATAGTGA GTGGGCCGCC	1260
	ATAAGCCATC ACTGGAATC CAACCCAGG GGTCCAGGAG TGATCTCTGA GTGACTCAAC	1320
55	AAAGACAGGA CACATGGGGT ACAAAGACAA GGCTTGACTG CTTCAAAGCT TCCCTGGACC	1380
	TGAAGCCAGA CAGGCAGAG GGTCCGCTG ACAAATCACT CCCATGATGA GACCTGGAG	1440
	GACTCCAAAT CCTCGCTGTG AACAGGACTG GACGGTTGCG CACAAACAAA CGCTGCCACC	1500
60	CTCCACTTCC CAACCCAGAA CTTGGAAAGA CATTAGCACA ACTTACGCAT TGGGGAATTG	1560

	TGTGTATTTT CTAGCACTTG TGTATGGAA AACCTGTATG GCAGTGATTT ATTATATAT	1620
5	TCTGTCCAA AGCCACACTG AAAACAGAGG CAGAGACATG TACTCTGGTG TGATCTCTTG	1680
	TCCTCAGTGT CTCTTCTGGG CTCTGTCCC TCTTGCTTTA TAGCTAGCTG CCCGGGGACC	1740
	AAGGTACAGG TGAAAGCAAG GTAGCAGCTT GCGGGAGGAG GCCTGTCTGG CTTACCAGTC	1800
10	TATACACTGT GGCCTCAACC TCCCAGACAG GGCAGAGAAC TGTGGGCAGC TCGTTTGCTT	1860
	TCTAGGCTGG CTGGAGAGGT GGGAGCTCAT TGATAGACTC ATGATGGAAA CTATTTTGA	1920
15	AACAGGCTTC CTCCTTCAGG AGAGATCATG CGGACTAAAC TGTAGCAATT CCAGTGCACC	1980
	TGGCAGTGAT CCTTTCTTTT GCAAAGTACT GTCTCTTTGG TTCCAGTAAG TTGGACCACC	2040
	ACATGACATY ATTTTCCCTG GAACCTGGTC ACTGACTAAC ACAGACAATT GGGACTCCAG	2100
20	AGCCTCAAGA GCCAGGAGAG GGCACAGTAC ATACAGAGGG AGTCAAATGG GATCTCATTT	2160
	TGAGTCCTGC CTTCGGCACA CTCAGAACGG CANCCCCAAG GCCCGGAGTG TCCAGGGCTT	2220
25	CTGGCCTGAG GTGAATCTGC CAGGCCCAAG AAGGCACAAA GGTAGGAGCA CAGAGAGCCC	2280
	CATTCCCACA GCGGKQGGC CCAGCAGCAC CAGTGAAGC TCAGCTGTCC TCCAGCTGCT	2340
	CTCGGCAGAC AGTTCAGTGC ACAGTTTATG CCTAGCTGA AAAAGATCTC CCGGACGTAT	2400
30	TTCAGCACAT CCTCTTCTC CTCCTCTCA GGGCTCTGC TACAGGCAGA GCTGGAACCC	2460
	CCCGCCTCT GGAAGGGCT GAGGCTTGA GYCAGTGCCT GTC	2503

35

(2) INFORMATION FOR SEQ ID NO: 96:

- 40 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2801 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

- 45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 96:

	CTGGAAAGCC GAGGGTAGCC GAGCGGGCG GCGCTCTGG AGCGGCGGT GCTCGGGCTG	60
50	CCGTCCGCTC CGCCAGAAGC ACCGAGCAGC CGAGCCGGG CCCGCCGCC TCCTCCTCCA	120
	TGAGGCCCGA GTGAGGCGCG GCGCTATAG CCGACCCCG GCGCCTTCCC CCCGCTCT	180
	ATCGCGAGCG CACGACMAGC GGGCCCTGGA GGAGGAGCG GAGGAGGAG AGCATGTCCG	240
55	ACGGTTTCGA TCGGGCCCCA GGTCTGTGTC GGGCCGGAR CCGGGGCTG GGCCGCGGAG	300
	GGGCGGGCC TRAGGCGGC GGTITYCGA AMGGARCGGR GCCTGCTGAG CGGRCGGGC	360
60	ACCAGCCGCC GCAACCCAAA GCGCCGGGCT TYCTGCARCC AMCGCCGCTG CGCCARCCA	420

	GGACGACCCC GCCGCCAGGG GCCCAGTGCG AGGTCCCCGC CAGCCCCCAG CGGCCTTCCC	480
	GGCCCCGGGC GCTCCCAGAG CAAACGAGGC CCCTGAGAGC TCCACCTAGT TCACAGGATA	540
5	AAATCCCACA GCAGAACTCG GAGTCAGCAA TGGCTAAGCC CCAGGTGGTT GTAGCTCCTG	600
	TATTAATGTC TAAGCTGTCT GTGAATGCCC CTGAATTTTA CCCTTCAGGT TATTCTTCCA	660
10	GTTACACAGA ATCCTATGAG GATGGTTGTG AGGATTATCC TACTCTATCA GAATATGTTT	720
	AGGATTTTTT GAATCATCTT ACAGAGCAGC CTGGCAGTTT TGAAACTGAA ATTGAACAGT	780
	TTGCAGAGAC CCTGAATGGT TGTGTTACAA CAGATGATGC TTTGCAAGAA CTTGTGGAAC	840
15	TCATCTATCA ACAGGCCACA TCTATCCCAA ATTTCTCTTA TATGGGAGCT CGCCTGTGTA	900
	ATTACCTGTC CCATCATCTG ACAATTAGCC CACAGAGTGG CAACTTCCGC CAATTGCTAC	960
20	TTCAAAGATG TCGGACTGAA TATGAAGTTA AAGATCAAGC TGCAAAAGGG GATGAAGTTA	1020
	CTCGAAAACG ATTTTCATGCA TTTGTACTCT TTCTGGGAGA ACTTTATCTT AACCTGGAGA	1080
	TCAAGGGAAC AAATGGACAG GTTACAAGAG CAGATATTCT TCAGGTGGT CTTGAGAAAT	1140
25	TGCTGAATGC CCTGTTTTCT AATCCTATGG ATGACAAATT AATTTGTGCA GTAAAATTGT	1200
	TAAAGTTGAC AGGATCAGTT TTGGAAGATG CTTGGAAGGA AAAAGGAAAG ATGGATATGG	1260
30	AAGAAATTAT TCAGAGAATT GAAAACGTG TCCTAGATGC AACTGCAGT AGAGATGTAA	1320
	AACAGATGCT CTTGAAGCTT GTAGAACTCC GGTCAAGTAA CTGGGGCAGA GTCCATGCAA	1380
	CTTCAACATA TAGAGAAGCA ACACCAGAAA ATGATCCTAA CTACTTTATG AATGAACCAA	1440
35	CATTTTATAC ATCTGATGGT GTTCCTTTCA CTGCAGCTGA TCCAGATTAC CAAGAGAAAT	1500
	ACCAAGAATT ACTTGAAAGA GAGGACTTTT TTCCAGATTA TGAAGAAAAT GGAACAGATT	1560
40	TATCCGGGGC TGGTGATCCA TACTTGATG ATATTGATGA TGAGATGGAC CCAGAGATAG	1620
	AAGAAGCTTA TGAAAAGTTT TGTGGAAT CAGAGCGTAA GCGAAAACAG TAAAGTTAAA	1680
	TTTCAGCATA TCAGTTTTAT AAAGCAGTTT AGGTATGGTG ATTTAGCAGA ACACAAGAGA	1740
45	GCAAGAAAAT GTGTCACATC TATACCAAAT TRAGGATGTT GAGTTATGTT ACTAATGTAT	1800
	GCAACTTTAA TTTTGTAA CACTATCTGC CAAAATAAAC TTTATTCCCT ATAACTTAAA	1860
50	ATGTTATAT ATATATAATA GTTTATTATG TACAGTTAAT TCTACTGTTT TGGCTGCAAT	1920
	AAAATCGATT TTGAAATAAA TGAAATGTTG AAAATTTTGC TAGTTGGTTA GATGCTTATC	1980
	CTTTAAATTC TACTTTTCTT GAGGGGAAAA AGTCTTCGTC TGGAAATACA TATTACTGCA	2040
55	AAAATGTAGC ATCCTTTTTT AGGTAGGAGT ATTATAGCTT YCATTTTAGT TKGACATTTA	2100
	GTGTCCCAAT GAATTGAATT TCAAAATGA ATCATAATCT TGAAATCTT TAGCACTAAA	2160
60	GTCTTGGGAA TATATCAACA ACTGATTTAC ATATGCAGAT GCTATTTGNA TACCAAGGGC	2220

	TTTTAAATG TCATGGGGG GAAAAACCCA ACTTGGTGA ACTCCAGCT AAACAACCAA	2280
	GACTTCACTG GAAGATTAT TCCAATTCTA GGAATTGTC TTTTATTATTT TTATTTTTC	2340
5	AACTGRCTAA CTCATTACC TTAAAGCCTA GAACATTATT CTGCTTTATT TATATGGCTT	2400
	TCTCACTTTT ATTTGTAGC AKGGGTGCA TCGACTTTT TACTAGAGAA TTTTACTAGA	2460
10	TATTTGTCAT TCAAGTTTC ATCTGCTTTA TAATTGATAC ACCTGAGGG TCACTTTTCT	2520
	AATACITTTA CTATAATGTG GTACCACCTC AGCCCTAATA AATAATATTT TTACCTAATG	2580
	TCAAATCTTT TTCCAGCTAA CTAAAACTG TGTACAAAAG GATGCTTGT AAATATGCAT	2640
15	GTAAATAGTT CTGTAATAA CCCACTGTTT TACATTGGT ACATCTGTGT CTGCTAATAC	2700
	AGTTAGCTTT CTCACITTTT TGCTTGTGTTT TTCAGTCTGA ATTAAAAATTA GACTTTGAAA	2760
20	ATAAGCTTA AAAAAAAAAA AAAAAAAAAA AAAAAGCTCGA G	2801

## (2) INFORMATION FOR SEQ ID NO: 97:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1631 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 97:

35	ATGGAGCCAA AGACAATCAC TGATGCTTTG GCTTCTAGTA TAATTAAGAG TGTGCTGCCT	60
	AATTTTCTTC CATACAATGT CATGCTCTAC AGTGATGCTC CAGTGAGTGA ACTGTCCTC	120
	GAGCTGCTTC TGCTTCAGGT TGCTTGCCA GCATTACTCG AACAGGGACA CACGAGGCAG	180
40	TGGCTGAAGG GGCTGGTGG AGCGTGGACT GTGACCGCCG GATACTTGCT GGATCTTCAT	240
	TCTTATTTAT TGGGAGACCA GGAAGAAAAT GAAAACAGTG CAAATCAACA AGTTAACAAT	300
45	AATCAGCATG CTCGAAATAA CAACGCTATT CCTGTGGTGG GAGAAGGCCT TCATGCAGCC	360
	CACCAAGCCA TACTCCAGCA GGGAGGGCCT GTTGGYTTTC AGCYTTACCG CCGACCTTTA	420
	AATTTTCCAC TCAGGATATT TCTGTTGATT GTCTTCATGT GTATAACATT ACTGATGCC	480
50	AGCCTCATCT GCCTTACTTT ACCAGTATTT GCTGGCCGTT GGTAAATGTC GTTTTGGACG	540
	GGGACTGCCA AAATCCATGA GCTCTACACA GCTGCTTG TGCTCTATGT TTGCTGGCTA	600
55	ACCATAAGGG CTGTGACGGT GATGGTGGCA TGGATGCCTC AGGGACGCAG AGTGATCTTC	660
	CAGAAGGTTA AAGAGTGGTC TCTCATGATC ATGAAGACTT TGATAGTTGC GGTGCTGTTG	720
	GCTGGAGTTG TCCCTCTCCT TCTGGGCTC CTGTTTGAGC TGGTCATTGT GGCTCCCTG	780
60	AGGGTCCCT TGGATCAGAC TCCTCTTTT TATCCATGGC AGGACTGGGC ACTTGGAGTC	840

	CTGCATGCCA AAATCATTGC AGCTATAACA TTGATGGGTC CTCAGTGGTG GTTGAAAAC	900
5	GTAATTGAAC AGGTTTACGC AAATGGCATC CGGAACATTG ACCTTCACTA TATTGTTCGT	960
	AAACTGGCAG CTCCTTGAT CTCTGTGCTG TTGCTTTCCC TGTGTGTACC TTATGTCATA	1020
	GCTTCTGGTG TTGTTCTTT ACTAGGTGTT ACTGCGGAAA TGCAAAACTT AGTCCATCGG	1080
10	CGGATTTATC CATTTTTACT GATGGTCGTG GTATTGATGG CAATTTTGTG CTTCCAAGTC	1140
	CGCCAGTTTA AGCGCCTTTA TGAACATATT AAAAATGACA AGTACCTTGT GGGTCAACGA	1200
15	CTCGTGAAC ACGAACGGAA ATCTGGCAAA CAAGGCTCAT CTCCACCACC TCCACAGTCA	1260
	TCCCAAGAAT AAAGTAGTTG TCTCAACAAC TTGACCTTCC CCTTTACATG TCCTTTTTTG	1320
	TGGACTTCTC TCTTTGGAGA TTTTCCAG TGATCTCTCA GCGTGTMTT TAAGTTAAAT	1380
20	GTATTTGACT TGTGTTCTCA GCATTCAGAG AGCAGCGGTG TAAGATTCTG CTGTTCTCCC	1440
	TGGATCTTCT GACATTACTG CTGCTGAGA TTGTATATG TGTAAATACA AGTTCCTTGA	1500
25	TACCCTAAAA CCTTGGATTA AACAGAATGT GCATTGTACA TCTTTAAACA AAATGTATAT	1560
	TAATTTATTA AATCTAGTTG TCACTTTAAA AAAAAAAAAA AAAAACTCG AGGGGGGCCC	1620
30	GGTACCCAAA T	1631

## (2) INFORMATION FOR SEQ ID NO: 98:

- 35 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 504 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - 40 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 98:

	CCGAGCTGGG CGAGAAGTAG GGGAGGGCAC GAGCCGCCG GGTGGCGGTT GCTATCGCTT	60
45	CGCAGAACCT ACTCAGGCAG CCAGCTGAGA AGAGTTGAGG GAAAGTGCTG CTGCTGGGTC	120
	TGCAGACGCG ATGGATAACG TGCAGCCGAA AATAAACAT CGCCCTTCT GCTTCAGTGT	180
50	GAAAGGCCAC GTGAAGATGC TGGGCTGGA TATTATCAAC TCACTGGTAA CAACAGTATT	240
	CATGCTCATC GTATCTGTGT TGGCACTGAT ACCAGAAACC ACAACATGA CAGTTGGTGG	300
	AGGGGTGTTT GCACTTGTGA CAGCAGTATG CTGTCTTGCC GACGGGGCCC TTATTTACCG	360
55	GAAGCTTCTG TTCAATCCCA CGGTCCTTA CCAGAAAAAG CCTGTGCATG AAAAAAAGA	420
	AGTTTGTAA TTTTATATTA CTTTTAGTT TGATACTAAG TATTAAACAT ATTTCTGTAT	480
60	TCTTCCAAAA AAAAAAAAAA AAAA	504

(2) INFORMATION FOR SEQ ID NO: 99:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1416 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 99:

5	GGCAGGAGGG AGGGAGCCCT CTCGGTTGGG TGACTCTTGT GTGCCCTTTA GACAGGCTGG	60
10	CCTGCCGGTT CCACAGGGTA CAGTTAGGAC TTGAGTCTTT CTTTTCTGT TTTGAGTTGG	120
	TGAGTGAGTG ATAGGGTAAC ATGGGCCCTC AGGATGACCC CTTGGAAC TGCCGAGTTC	180
20	CTTAAATCTC AGCTGGGATC CTGGACCTGG GAGGCCCTG TGAGGGCCAG CTCTGGAAAA	240
	ACCTGGGAGT TGATGCCGA GCTGTGAAG AACTCTGCTC GAGGGCAGG TGCCCTGGAA	300
25	CACTGGTAGT TCTGGGGCTG GGAGGGAGAG GGGCTCCGGC TTTCTCTGAA ATGAACACTG	360
	CTCTTCAGCA GTTCAAGTAC TTGTTCTCAA AACATTTTCT AATTGATTGG TAGGTTTCA	420
	TAAGCATGTG TTCTTTAAGG CATGGAAGG GAAGAATGCT CAAGCAAGTC ATGTTTGTMT	480
30	TCAGTGGGAT GGGCCCGCGT TCTCACTGCT GGGGGCTTCC CCTTCATGTG GCACCTTTGT	540
	GCAGGGGCCA CCAGGCAGAC TCTTCCCACC TTCTCCCCT GAAGCACCAA GGGGCTTGA	600
35	ACCGTAATTT GGCTAATCAG AGGCAITTTT TTGTCTTAG TATCTTTCAC ACTTGTCCAA	660
	CCGTCTTATT TTTTAAAG TTCTGTGCT TGTATTAACA CGAACTAGA GAGAAATAGT	720
	TTCTGAAGCC AGTTTATTGT GAAGATCCCC AAGGGAGGT TCGGTAGAGA AAAATAGTAA	780
40	GCTGGTTTAA AACTGACGA GGGCAAACAG CCAGGACGCA TTGGAGAGGA ATTTGCCAAA	840
	GATCTACCCCT GAGATAACGC CTGTCCAGTG TCTTACCAC GTGAATAACC AGCGCTCCAA	900
45	AGTGTTTTTC TGCTTTGAAA AAAAAATTC CACAAGCTTT TAAAGGTGCA TTTAAGAATC	960
	CATGTGACTT TAGAATGGAA CTGCCGGCCC TGGCAACTGT CACGTGTGCT AGAAGGTTGG	1020
	ATGCCTCTGG AATGCATGTG ATACTCATCT CCATTTTGT TCTTGATTG CATTTTGT	1080
50	CTTTTAGCAG ATCTGTCCCT GTGGGTGGTG TCTAAGAAGT CGGACACCTT GGTTTTGTG	1140
	TTAGATTGAG CTGGGCAGCT GCAATCAGCT TCTTTATATG CAAATTAGGC ACGACCCATC	1200
55	TGTGGTTTCT GGTGGTGGC TAATGAAGTG AGGGGAGGGA GGGATGTCAC CCCAAAAGTA	1260
	GGCCCTCCCA TTGGCTTTGG CCAGGCCAGA CACTTCACAT CGTTTACATG GTTCTGTGTA	1320
	ATTTTAAAGT TTATGTGTAT AAAGCGAAGC TGTTCGTG TGAACTGTATA TTTGTAAAT	1380
60	AAATATATTG CTACTTGAAA AAAAAAAAAA AAAAAA	1416

## 5 (2) INFORMATION FOR SEQ ID NO: 100:

## (i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 2847 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 100:

15 GGCTAGGACA ATTTTGGTGC TTACCTATC TCTGCAAAGA CTGGAGAATT TGGCATACCA 60  
TTAATTACAA CCACCAATCA TATCCAACAA AAGTACCTA AAAGAAGGAC CAGTGGCCAC 120  
TCTCGAAAAA ATTAAAGTAT CAGAAGATTA AAAAGATTTT AGGATTGGA AGCTTGTATT 180  
20 GTCTTTCCCC AATAATCATT GTTTGATCTC CAAATAGTAG CCTTATATTA GCAATRGACA 240  
GATCATTTGGT TCTCCATATC TGATCATATG TTACTACTTT GGAATCAGTA TTTGGGCAAA 300  
25 TTCAAGCATT TATGCAGTGG ATATAAATGG AAATATAAAA ATATTTGCCA ACCTGTCTCA 360  
GTAACCTATC ATATCTCTGT GNATCCTCAA GGAAAGCACT TTTGCTTTTA CTTAGAAAGC 420  
30 GTTTCAGATT TGCTTTATAG ACTCCTGCTG TCTTCAGTAC CTGATAAAAC TTTAACCAGG 480  
GAAGCATTAA ACACAGTGCA GCAGCTTTTG CCCAGGCTTC TAAGTTCCTG CCGGCAGCAT 540  
TTATCAATGT AAGAACTAGG ATGCTTCCTG CAGTGGCACT ACCTTCCCCT AGAGCTGGAG 600  
35 CATGCTGCTT GGCCTTAAGC CCCAGCATGA TGAGGCTTCC CTCCTGCCAG GTCAGTAAAA 660  
GTTAGAGAGC TCAGAATTGG GTCTTGCTG GGTGCAGGTG GCAGGGTTTG CTGAAACCCC 720  
TAAAGAGAAG TCACCAAGGG AGGCAGGTAA TGAATGTTT CAGAATCAGT CKGATACTCA 780  
40 TAGCAATTC TGGCTATCTT TCAAATGTTG AATTTCTGGA TGCTGAGAGG GACTTTGATT 840  
TGATATCAAT AAATCCAGGA CAGTCCCAAG AAGTGCTTGG AGTCTCGGCT CTGACAGCCC 900  
45 AAGAAGGGAA ATAACCTGTA TTAAGGAACA ACTATGAGCC AGGCCCTGAG CTGTCTCTTA 960  
GATAATAAAA CAGATGGGGA GTGGAAGAGT CATTTGCTTC AAGTTATACA GCTAGGAAAT 1020  
ACTCAAGCCA AATCTTGAAC GCAGCTCCCC CTAATCTGT GGACAGGCAC TTTGTACCAC 1080  
50 ACACCATGGT CCACCTAAAA ACAGAAGGAT AAAAAGACTT CAGGTTTTC CACTGTGTGC 1140  
TGACCATCCC AATTTATGAA TCTTCTTCAA AATGACATTT CACAGTTATA GTTAGGGCTC 1200  
55 AGAAATGGCA TTGAGGTAGC CTTATTTCTC CCCTTAGCA GATGCTTTAA GTACACATTG 1260  
CTGACTTGAG CCCACCCCCA GGAGTTAGGA GAACATTTCC TTTTTCATGC CATCTTCCAT 1320  
60 AAATAAGGTG TTTCTTGGCC TTCAAAGATA TAGAACTTTG CAGCAGTAGT AAAAGTGAAG 1380



	GGTGTCTGTC TCTCTACTCA ACTTTATTTG AAAATGTCTG CAGCTTCACT CCTGTAGAAA	1440
	AGGAAATCTT CATATTTTAG TAAACTTAGC CGCCAGTGTA CTCTGTGAGG ATGTGGCAAT	1500
5	TCAAAGTCCA GTGAATCTGG CTCTCTTACT GATTCCTGGT TTTAGTGTGT GTGTGGGGG	1560
	AGTGTGTACC TATATATAAA GGACAAGTGT GATATGTGTG TATATGTATA TACATACATA	1620
10	CATGTCCACA CACACACACA CAATATTTGA GAGCTAAGGA AAACCTCAAAG CAGCCCCCTC	1680
	ATTATCTTGC GTACTACTTC AAAGATTTCT GTCAGCCCTA ATTACAAGTG TCACCATATA	1740
	GTGGGGCTT AGGTACTTGC TTACAGGAAG AGCAATTCCC TAGCAAAGGT CATTAGCTCC	1800
15	TAAGGCACTG AGTCAAAGTG ACAGCCCTGA AGGAAATGTC ACTCCAGCCC TCCTCCAGGA	1860
	TGTCTAATAA GATGGGAAAC TTGGATGCCC AGCCATTTTG GTGACCTGAG AGTCTAACTA	1920
20	CTCCAGTTAG ACCTAAGGGC ACAAATGCAG AATTCATGAC CTTGTAGTTG TGGCAGGGTC	1980
	TAGGAAGTCC TCTCTCCCCA AGTAGAAAAT ATTCTCTTGC CATTCTTGAA ATTCCACATT	2040
	CATATAATGG CTGTGCAATA CATGCTTCTC AATAAGAAAA TTAACATGAT GTTTACTGTG	2100
25	TGCTGATCAC ATCAGATTTT TATGTTTAA AAAATCTCAT TATGNTTGA GTCCAGCCCA	2160
	GCTCTAAGAG AAAAGAAGG CCCATATGGG AGACTTCAGT CTCATTATTA TTGCCCTTAT	2220
30	CCAGCAGTGC TTATRAAGCC CCCTACCCTG TCCCATPCCA GAAACCATAA GACTCAGGCA	2280
	GTCTTGATT CTGGAGGCCT GCCTGGTAAG ATAAGATAGT ATAATTTGGA ACTGAGAACA	2340
	TACCAGAAAC AGCAGAACGA GGGCCAGAGC AGAAAAATGA AAATAAGTGG AGACACTTAT	2400
35	GGATACATTG GTGCAAAAAA AGCCACGGGS CCCATACTGG GCTTGATATG ACTTTGAGGG	2460
	GACAGCAGAT TAATACTTAA TGAGGGTTAA ACCTGACCAG TCTTTCTACA GTGACAGGCC	2520
40	AACTGCGATG AATGGGGAGA ACCAATGAAT CCATTGTCCT CTGCCTATTT TCCTGTGCAC	2580
	AGTCACATTC CCTCCTTAGG AATCTTCCCC TTCCACCCCT TACATTAAAC AAGGGAACAC	2640
	TGAATCTTTC AAGGGAATTA CACGTTTGGG TTAATGTTTC AGTATATCAT TTTCTACTG	2700
45	TAAATTATTT TGTAAGAGAG ATTTACTGCT ATCCCAGGAT GTTCGGACTT GGTGCCCCTG	2760
	TGCATTTGGA AATCAATAAA CTATTACTGG AAATGCCAAA AAAAAAAAAA AAAAAAAAAAN	2820
50	NAAAAAATC GAGGGGGGCC CGTACCC	2847

(2) INFORMATION FOR SEQ ID NO: 101:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1394 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 101:

5	GAGATTGGTG GAGGAGAGTA AATAATCTAG AGGCAAGAGT TCAGTGAGGG CCAAGGGGGA	60
	CCCCCAGAAA AAGGTATGGA GCTAACTCAT CTCCTTTTACA AGGGGTGGCC ATGACTTACT	120
	GTTCGAAAGT ACTCAGTGTA TATTTAATGT TGATTGTTGA ATTTTAGTTA CGAGAGGGAA	180
10	GAACAATTTT ACTTCTGTCC TTATTTCACT TGCTGAAAAG CTGTGGGACA AAATGTATGG	240
	AATAGACAAG GCCACTTTCT TTGTGATTC TGCTTTTCAT GCATATTATT TTATTTACCC	300
15	ATAATTTCCA AGAGGTTTGG CGTTCGCTC TCCTGCTTTT TTCTTTTCATC CACCCCTTTC	360
	CTTTTTTTTG AAGGGGGTTA TATATGAGAG TTCATTGAAG AAGTCCAGTG AGGCTGAAGT	420
	AAAGGGGCAA GATAGGGCAG TTAAGTAAAG AGCACTTTAT TTCTTTGAAG CCTTTCTAAG	480
20	AAAGAAATGG GGTGCGAGT GGCTTGAATC TCCCATGATG TTGGAGGGCA CTTAGTGGGG	540
	TTGAAGTATG ACATAATATT TOCCATGGG GAAAGGAGAA TTCTCTTAG AGGGTGGCAA	600
25	AATGCCCTTG CCCAGTGTC CTATTTTAGG CATCTTTTCC TTCCCTATTC CTCCAGTCA	660
	GGGTGTGTCC TATACAAAAC TTCCCATCAG TTCTCCTCAA TATTCCTCAT TTGTAAATGA	720
	TCACTTCTCT TTCTTAAACC CTTTTCCTGT TCAGATCCAT ACAGGATTG CAAGGGTAGG	780
30	ATCATACATG CAAATGCCCC TTGTTTCATCT GTGTCTCTG CAACTAGTC TCATGAAGAA	840
	TTCTGGCGTG CAGCAGGGTA GCTGAAGTTT GGGTCTGGGA CTGGAGATTG GCCATTAGGC	900
35	NTCNCTGAGA TTCCAGTCC CTTCACCAA GCCCAGTCTT GCTACGTGGC ACAGGGCAAA	960
	CCTGACTCCC TTGGGGCTC AGTTTCCCCT CCCCTTCATG AAATGAAAAG AATACTACTT	1020
	TTCTTGTG GTCTAGCATT GCTGGACACA AAGTGTAGTC ATTATTGTG TATTGGGTGA	1080
40	TGTGTGCAAA ACTGCAGAAG CTCACTGCCT ATAAGAGGAA ATAAGAGAGA AAGTGGAGGA	1140
	GAGGGACAAA AGGAGTAATT ATTTGGTATA GATCCACCCA TCCCAACCTT TCTCTCCTCA	1200
45	GTCCCTGCTC CTCATGTTTC TGGTTTGGTG AGTCCCTTGT GCCACCACCC ATAATGCTTT	1260
	GCATTGCTGC ATCCTGGGAA GGGGTATAT GGTCTCACA GTTGTGTCA TTGTTTTTTT	1320
	GCATGCTTTC TTAATAAAAA AAAAAAAAAA ATGTTTANAG TTTTATCTTA AAAAAAAAAA	1380
50	AAAAAAAAA ACCC	1394

## 55 (2) INFORMATION FOR SEQ ID NO: 102:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 794 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

60

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 102:

5 GGMRCGAGGC GGAGTAAAGG GACTTGAGCG AGCCAGTTGC CGGATTATTC TATTTCCCTT 60  
CCCTCTCTCC CGCCCCGTAT CTCTTTTCAC CCTTCTCCCA CCTCGCTCG CGTACCATGG 120  
CGGAGCGTCG GCGGCCACTC AGTCCCATTC CATCTCCTCG TCGTCCTTCG GAGCCGAGCC 180  
10 GTCCGCGCCC GCGCGGGCGG GGAGCCCAGG AGCCTGCCCC GCCCTGGGGA CGAAGAGCTG 240  
CAGCTCCTCC TGTGCGGTGC ACGATCTGAT TTTCTGGAGA GATGTGAAGA AGACTGGGTT 300  
15 TGTCTTTTGA CACGCTGATC ATGCTGCTTT CCTTGGCAGC TTTCAGTGTC ATCARTGTGG 360  
GTTTCTTAMC TCATCCTGGC TCTTCTCTCT GTCACCATCA RCTTCAGGAT CTACAAGTCC 420  
GTCATCCAAG CTGTWCAGAA RTCAGAARAA GGCCATCCAW TCCAAAGCCT ACCTGGACGT 480  
20 AGACATTACT CTGTCTCAG AAGCTTTCCA TAATTACATG AATGCTGCCA TGGTGCACAT 540  
CAACAGGGCC CTGAAACTCA TTATTCGTCT CTTTCTGGTA GAAGATCTGG TTGACTCCTT 600  
25 GAAGCTGGCT GTCTTCATGT GGCTGATGAC CTATGTTGGT GCTGTTTTTA ACGGAATCAC 660  
CCTTCTAATT CTGTCTGAAC TGCTCATTTT CAGTGTCCCG ATTGTCTATG AGAAGTACAA 720  
GACCCAGATT GATCACTATG TTGGCATCGC CCGAGATCAG ACCAAGTCAA TTGTTGAAAA 780  
30 GATCCCAAGC AAAA 794

35

(2) INFORMATION FOR SEQ ID NO: 103:

(i) SEQUENCE CHARACTERISTICS:

40

- (A) LENGTH: 1544 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 103:

TTTGCTTGCT AGTCTGAACC AAAGAGTTGT TTGGGCATTT GCTGTGTTGG CCATTTCTGG 60  
AGCAAGAGGG TCTTCTTCCT CCTTCCCCCA GCCAGCCAGC TGTCTGGGG CCAGGCTTTC 120  
50 CTGGGTGGAA AGAAGTATAC CTTTCCCTGG GGGCCTAGGA TAGCAAAGTG AGCCATAGTG 180  
GGCAGGCTG CCTCCATGC TGGGCCCCAG CCCAGGTCTG CACTCGCCTG GATCACCTTC 240  
TTTGAGCCTT AGCCATCTCC TGTAGGTAG GAATGAACCT GCCAGCCTTC AGGYTCGTTT 300  
55 AGCTATGACC ATCTGTGCGG TCAGGGTACA CTCAGCTCTC CTCCCCAACT CCAGCAGCCT 360  
TTAAGAAGTG TCCCTTTGGC GCCCCCTGGA GGCAGAGCAC TGAGCTGGAC CCTGGGTAGA 420  
60 CTCCACAGG GAGGACGGAG CTGGCCTCAG GAGTGGACA CCCAGACTTG GCAGGGCCTT 480

	CAAGAGGCCT GTGTGGGGGC CCCAGGAATC CTTAGCTGAA GCGGGGAGAC TCACTCTCCA	540
5	TCTCAGGAAA TTCTAGCCCT TGCCCTCAGG GAGCCACGGT TGAGGGTGAG GCCCAACACC	600
	TGCCTTAGGG CCCTGGGTGG GCAAGTCTGG GCCCTGGGGT AGGGAGGGAG ACTCAGGCCC	660
	ACACTTGGGT ATTTTCTAAT TTCAGACAAA CACACACTCA GCGCGCACTC ACTGATTCTT	720
10	ACACATTGCC AAGATTTTAC ACATGTGACC AGGGGCCACC AAAGTCCCTG TGACCTTTGT	780
	GACTAGGATC CTAATTTCTC TATTTTCTCC TGGGTGCCTG GGTCTGTGTC ACCTGGGGCA	840
15	GTGTGGATAA TGTTTAGTTC TGTGACACTG TTTTTTGGGG GTGGCACCTG GTTCTCCGAT	900
	GCCTGGGCTG GTGTCAAGGC CAGGACTGTA GTGCTGGGAG CAGTAAAGCT CAGCTCTGTG	960
	TAATGAGTGA TGCTATGGCT TGCTCGTGTG TTATGATCCA ATCCTTTTCT ACATCAGCCC	1020
20	TTGTTTGTGTT TTATGGCTAG TCTTATCTGG CCTGGTTATT TCCTTGCGGG GAGGAGAGGG	1080
	TTTGCTAATC TGCTCCAGC CCAACCTATT ACCACCCAC CTCGCTGGGA CCTACTGCTC	1140
25	GGGAGGCAGC AGACAGGGAG CCACCAGCAG TGGCTTCCTG GCCCTGTGCT GGGGGTGGG	1200
	GGAAGCTGGG GGCACATGTG GCCCTTGCCT TCTGAGCAGC TCCAGTGCC AGGGCTTTGA	1260
	GACTTTCCCA CATGATAAAA GAAAAGGGAG GTACAGAAGT TCCAATTCCC TTTTATTTT	1320
30	GCTGGTGGT ATCTGTAAAT GTTTAATAAA TATCTGAGCA TGTATCTATC AACGCCAAGA	1380
	ATTTCAAAGT CTCCTTCAAC AATATGAGGC TTTTAGGATG TTTATATTCC TTCATCCCTC	1440
35	TTGTTTCCCA GGTTTTGCAG GGAAGGAGG TCTGGAATTA TAGATACAGC TTATTATTAA	1500
	ATTTGTCTTT GCATAAAAAA AAAAAAAAAA AACNCNNGGG GGGG	1544

40

(2) INFORMATION FOR SEQ ID NO: 104:

(i) SEQUENCE CHARACTERISTICS:

45

- (A) LENGTH: 871 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 104:

	ACCCACGCGT CCGNCTTGTC CACCCGGGGG CGTGGGAGTG AGGTACCAGA TTCAGCCCAT	60
	TTGGCCCCGA CGCCTCTGTT CTCGGAATCC GGTGCTGCG GATGAGGTC CCGTTCTTA	120
55	AGGTGGGTG CTGTCCACCC GGGGGCGTGG GAGTGAGGTA CCAGATTGAG CCCATTGGC	180
	CCCAGCCT CTGTTCTCGG AATCCGGGTG CTGCGGATTG AGGTCCCGGT TCCTAACGGA	240
60	CTGCAAGATG GAGGAAGGCG GGAACCTAGG AGGCCTGATT AAGATGGTCC ATCTACTGGT	300

CTTGTCAGGT GCCTGGGGCA TGCAATGTG GGTGACCTTC GTCTCAGGCT TTCCTGCTTT 360  
TCCGAAGCCT TCCCCGACAT ACCTTCGGAC TAGTGCAGAG CAAACTCTTC CCCTTCTACT 420  
5 TCCACATCTC CATGGGCTGT GCCTTCATCA ACCTCTGCAT CTTGGCTTCA CAGCATGCTT 480  
GGGCTCAGCT CACATTCTGG GAGGCCAGCC AGCTTTACCT GCTGTTCTTG AGCCTTACGC 540  
10 TGGCCACTGT CAACGCCCCG TGGCTGGAAC CCGCACCAC AGCTGCCATG TGGGCCCTGC 600  
AAACCGTGGG AGAAGGAGCG AGGCTGGGT GGGGAGGTAC CAGGCAGCCA ACAGGTTCCC 660  
GATCCTTAAC GCCAGTTCG AGAGAAGGAC CCAAGTACA GTGCTCTCCG CCAGAATTTT 720  
15 TTCCGCTACC ATGGGCTGTC CTCTCTTTC AATCTGGGCT GCGTCCTGAG CAATGGGCTC 780  
TGTCCTCGTG GCCTTGCCCT GGAAATAAGG AGCCTCTAGC ATGGGCCCTG CATGCTAATA 840  
AATGCTTCTT CAGAAAAAAA AAAAAAAAAA A 871  
20

25 (2) INFORMATION FOR SEQ ID NO: 105:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 404 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 105:

GGCAGGAGTT ATAGCATGGC ATTCATACTT TTGTTTTATT GCCTCATGAC TTTTITGAGT 60  
35 TTAGAACAAA ACAGTGCAAC CGTAGAGCCT TCTTCCCATG AAATTTTGCA TCTGCTCCAA 120  
AACTGCTTTG AGTTACTCAG AACTTCAACC TCCAATGCA CTGAAGGCAT TCCTTGTCAA 180  
40 AGATACCAGA ATGGGTTACA CATTTAACCT GGCAACATT GAAGAACTCT TAATGTTTTC 240  
TTTTTAATAA GAATGACGCC CCACTTTGGG GACTAAAATT GTGCTATTGC CGAGAAGCAG 300  
TCTAAAAATT ATTTTTTTAA AAAGAGAAAC TGCCCATTA TTTTGGTGGG GTTGGTTTTT 360  
45 AATTTNTAAT NTGAAAAATT TTTTGGGGT TTTTGGGGCC ATGG 404

50

(2) INFORMATION FOR SEQ ID NO: 106:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1542 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 106:

	GTCAGACAGG TGGAGCCGCC GGGGCAGGAG TCTCAAAGAG CCAGGCTCCA GGAGAGGAAG	60
	GGCTCTRCGA GAGGAGAGAG GAGAGCGCTG GAGAGGAGAG GCTGGAGAGT CCTTAGCCAG	120
5	GATGGAGGCT GTTGTGAACT TGTACCAAGA GGTGATGAAG CACGCAGATC CCCGGATCCA	180
	GGGCTACCCT CTGATGGGGT CCCCTTGCT AATGACCTCC ATTCTCCTGA CCTACGTGTA	240
10	CTTCGTTCTC TCACTTGGGC CTCGCATCAT GGCTAATCGG AAGCCCTTCC AGCTCCGTGG	300
	CTTCATGATT GTCTACAACT TCTCACTGGT GGCACCTCTCC CTCTACATTG TCTATGAGTT	360
	CCTGATGTGG GGCTGGCTGA GCACCTATAC CTGGCGCTGT GACCCCTGTGG ACTATTCCAA	420
15	CAGCCCTGAG GCACTTAGGA TGGTTCGGGT GGCCTGGCTC TTCTCTTTCT CCAAGTTCAT	480
	TGAGCTGATG GACACAGTGA TCTTTATTCT CCGAAAGAAA GACGGGCAGG TGACCTTCCT	540
	ACATGTCTTC CATCACTCTG TGCTTCCTTG GAGCTGGTGG TGGGGGGTAA AGATTGCCCC	600
20	GGGAGGAATG GGCTCTTTCC ATGCCATGAT AAACCTCTCC GTGCATGTCA TAATGTACCT	660
	GTACTACGGA TTATCTGCCT TTGGCCCTGT GGCACAACCC TACCTTTGGT GGAAAAAGCA	720
25	CATGACAGCC ATTCAGCTGA TCCAGTTTGT CTTGGTCTCA CTGCACATCT CCCAGTACTA	780
	CTTTATGTCC AGCTGTAACCT ACCAGTACCC AGTCATTATT CACCTCATCT GGATGTATGG	840
30	CACCATCTTC TTCATGCTGT TCTCCAACCT CTGGTATCAC TCTTATACCA AGGGCAAGCG	900
	GCTGCCCCGT GCACTTCAGC AAAATGGAGC TCCAGGTATT GCCAAGGTCA AGGCCAACTG	960
	AGAAGCATGG CCTAGATAGG CCCCCACCTA AGTGCCTCAG GACTGCACCT TAGGGCAGTG	1020
35	TCCGTCAGTG CCCTCTCCAC CTACACCTGT GACCAAGGCT TATGTGGTCA GGA CTGAGCA	1080
	GGGACTGGC CCTCCCTCC CCACAGCTGC TCTACAGGA CCACGGCTTT GGTTCCTCAC	1140
40	CCACTTCCCC CGGGCAGCTC CAGGGATGTG GCCTCATTCG TGTCTGCCAC TCCAGAGCTG	1200
	GGGGCTAAAA GGGCTGTACA GTTATTTCCC CCTCCCTGCC TTAAAACTTG GGAGAGGAGC	1260
	ACTCAGGGCT GGGCCACAA AGGGTCTCGT GGCCTTTTTC CTCACACAGA AGAGGTCAGC	1320
45	AATAATGTCA CTGTGGACCC AGTCTCACTC CTCACCCCA CACACTGAAG CAGTAGCTTC	1380
	TGGGCCAAAG GTCAGGGTGG GCGGGGCTT GGAATACAG CCTGTGGAGG CTGCTTACTC	1440
50	AACTTGTGTC TTAATTAAAA GTGACAGAGG AAACCANAAA AAAAAAAAAA AAAAACTCGA	1500
	GGGGGGCCCG TACCCAAATC GCCGGTATGA TCGTAAACAA TC	1542

55

(2) INFORMATION FOR SEQ ID NO: 107:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2327 base pairs

(B) TYPE: nucleic acid

60

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 107:

5	GGTAGCTCAN TGCAGTGAAA TAGTCTTACT GGAAACAAAG CCTTTTATCA AGAATAATTA	60
	ACTCTTCCCT TTCTTTTTTG GAGAGGTGCT TTGTTTCTGA TCGGACCATT TCACTGCAGC	120
10	AAGCAACACA GTATTCTRAG CAGAAGATCG GGACTTGAGG CCATGTTGCG GAGGGCCAGT	180
	RACATTATCT GGACTCTGGA GTGTGAGGAA TATGGACTCC ACTCTTCACT ATATTACAR	240
15	CGATTTCAGAC TTGAGCAACA ATAGCAGTTT TAGCCCTGAT GAGGAAAGGA GAACTAAAGT	300
	ACAAGATGTT GTACCTCAGG CGTTGTTAGA TCAGTATTTA TCTATGACTG ACCCTTCTCG	360
	TGCACAGACG GTTGACACTG AAATTGCTAA GCACTGTGCA TATAGCCTCC CTGGTGTGGC	420
20	CTTGACACTC GGAAGACAGA ATTGGCACTG CCTGAGAGAG ACGTATGRGA CTYTGGCCTC	480
	AGACATGCAG TGGAAAGTTC GACCGAACTC TAGCATTCTC CATCCACGRG CTTCAGTTA	540
25	TTCTTGGAGA TCAATTGACA GCTGCAGATC TGGTTCCAAT TTTTAATGGA TTTTAAAAG	600
	ACCTCGATGA AGTCAGGATA GGTGTTCTTA AACACTTGCA TGATTTTCTG AAGCTTCTTC	660
	ATATTGACAA AAGAAGAGAA TATCTTTATC AACCTCAGGA GTTTTGGTG ACAGATAATA	720
30	GTAGAAATTG GCGGTTTCGA GCTGAACTGG CTGAACAGCT GATTTTACTT CTAGAGTTAT	780
	ATAGTCCCAG AGATGTTTAT GACTATTTAC GTCCCATGTC TCTGAATCTG TGTGCAGACA	840
35	AAGTTTCTTC TGTTCGTTGG ATTTCTTACA AGTTGGTCAG CGAGATGGTG AAGAAGCTGC	900
	ACGCGGCAAC ACCACCAACG TTCGGAGTGG ACCTCATCAA TGAGCTTGTG GAGAACTTTG	960
	GCAGATGTCC CAAGTGGTCT GGTCCGCAAG CCTTTGTCTT TGTCTGCCAG ACTGTCATTG	1020
40	AGGATGACTG CCTTCCCATG GACCAGTTTG CTGTGCATCT CATGCCGCAT CTGCTAACCT	1080
	TAGCAAATGA CAGGGTTCCT AACGTGCGAG TGCTGCTTGC AAAGACATTA AGACAACTC	1140
45	TACTAGAAAA AGACTATTTC TTGGCCTCTG CCAGCTGCCA CCAGGAGGCT GTGGAGCAGA	1200
	CCATCATGGC TCTTCAGATG GACCGTGACA GCGATGTCAA GTATTTTGCA AGCATCCACC	1260
	CTGCCAGTAC CAAAATCTCC GAAGATGCCA TGAGCACAGC GTCTCAACC TACTAGAAGG	1320
50	CTTGAATCTC GGTGTCTTTC CTGCTTCCAT GAGAGCCGAG GTTCAGTGG CATTCGCCAC	1380
	GCATGTGACC TGGGATAGCT TTCGGGGGAG GAGAGACCTT CCTCTCCTGC GGACTTCATT	1440
55	GCAGGTGCAA GTTGCCCTACA CCCAATACCA GGGATTTCOA GAGTCAAGAG AAAGTACAGT	1500
	AAACACTATT ATCTTATCTT GACTTTAAKG KKWAWKMMW KCTCAGMSRA TTATAMITSW	1560
	CWMRARGSM WYMAAWSCTK SWGCTCYWCC KSRSTGRMKG MMRCTCTAGA AYTRGYRGAK	1620
60	CMYYKSGCT KMWGGAAKKS GGCASGAGCC AGAGACCTGC ATTGCTTTCT CTGGTTTTA	1680

5 TTTAACAATC GACAAATGAA ATTCTTACAG CCTGAAGCCA GACGTGTGCC CAGATGTGAA 1740  
 AGAGACCTTC AGTATCAGCC CTAACCTCTC TCTCCCAGGA AGGACTTGCT GGGCTCTGTG 1800  
 GCCAGCTGTC CAGCCCAGCC CTGTGTGTGA ATCGTTTGTG ACGTGTGCAA ATGGGAAAGG 1860  
 AGGGGTTTTT ACATCTCCTA AAGGACCTGA TGCCAACACA AGTAGGATTG ACTTAAACTC 1920  
 10 TTAAGCGCAG CATATTGCTG TACACATTTA CAGAATGGTT GCTGAGTGTC TGTGTCTGAT 1980  
 TTTTTCATGC TGGTCATGAC CTGAAGGAAA TTTATTAGAC GTATAATGTA TGTCTGGTGT 2040  
 TTTTAACTTG ATCATGATCA GCTCTGAGGT GCAACTTCTT CACATACTGT ACATACCTGT 2100  
 15 GACCACTCTT GGGAGTGCTG CAGTCTTTAA TCATGCTGTT TAAACTGTIG TGGCACAAGT 2160  
 TCTCTGTGCC AAATAAAATT TATTAATAAG ATCTATAGAG AGAGATATAT ACACCTTTGA 2220  
 20 TTGTTTCTTA GATGTCTACC AATAAATGCA ATTTGTGACC TGTAAAAAAA AAAWAAAAAA 2280  
 ACTCGAGGGG GGCCCGGTAC CCAAATCGCC GATATGATCT AANCATC 2327

25

(2) INFORMATION FOR SEQ ID NO: 108:

30 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1062 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 108:

GGCCGCGAG GCGCAACAGC CGTCTGTCA GCTCTGGGTC CAACCGGACT AGCGAANATC 60  
 40 TTCTCATCC TCATCATCGT CTCTCTCATC CCGATCTCGG TCCAGGTCCC TCTCCCCCCC 120  
 ACACAAGAGG TGGCGAAGGT CCAGCTGTAG TTCTCTGGA CGTCTCGAA GATGCTCTTC 180  
 CTCTCTTCG TCATCATCTT CCTCTTCGTC TTCTCATCC TCATCATCCA GTTCTCGAAG 240  
 45 CCGCTCACGA ATCCCATCC CCGCGCCGA GRAAGTGACA GGAGGCGCG GTACAGCTCT 300  
 TATCGTTCAC ATGACCATTA CCAAAGGCAA AGAGTGCTAC AAAAGGAGCG TGCAATAGAA 360  
 GAAAGAAGGG TGGTCTTCAT TGGAAAGATA CCTGGCCGCA TGA CTGATC AGAGCTGAAA 420  
 50 CAGAGGTTCT CCGTTTTTGG AGAGATTGAG GAGTGCACCA TCCACTTCCG TGTCCAAGGG 480  
 GACAACTACG GCTTCGTAC TTATCGCTAT GCTGAGGAGG CATTTGCAGC CATTGAGAGT 540  
 55 GGCCACAAGC TGCGGCAGGC AGATGAGCAG CCCTTTGATC TCTGCTTTGG GGGCCGAAGG 600  
 SWGTNCTGCA AGAGGAGCTA TTCTGATCTT GACTCCAACC GGAAGACTT TGACCCAGCA 660  
 60 CCTGTAAAGA GCAAATTTGA TTCTCTTGAC TTTGACACAT TGTGAAACA GGCCAGAAG 720



AACCTCAGGA GGTAACCTTG GGCCCTTCCC TGCTATCCTT TTTCTCCTTT GGAGGTGCCC 780  
 AACCTCCTCC ACCCCCTTCC CTTACTCTAG GGGAGAGAGC TGCTAGTGAG ATGACTGTTT 840  
 5 TATAAAGAAA TGAAGAAAAG TGAAATAAAA AATATGTTGA ATCAGATTTT TTAAGAGGG 900  
 TATTGTTTTT TTTATAACAG GTATTGAAAC AAGTTAACTT GCATTCCTAT GTAAGATAGG 960  
 10 AGGGGCTGAG GGGATCCCCA GTGTTTGGA CATAAGTCAC TATGCAGACT AATAACATC 1020  
 AACTAGAGAG NAAAAAAAAA AAAAAAAAAA ATTTAAAAAA CT 1062

15

(2) INFORMATION FOR SEQ ID NO: 109:

(i) SEQUENCE CHARACTERISTICS:

20

- (A) LENGTH: 2539 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 109:

GAGAGACTCA CACTTCTTTT CCATTATCAC TGACGATGTA GTGGACATAG CAGGGGAAGA 60  
 GCACCTACCT GTGTGTTGTA GGTGTTGTTGA TGAATCTCAT AACCTAAGAG AGGAATTTAT 120  
 30 AGGCTTCTCTG CCTATGAAG CCGATGCAGA AATTTTGGCT GTGAAATTTC AACTATGAT 180  
 AACTGAGAAG TGGGGATTAA ATATGGAGTA TTGTCGTGGC CAGGCTTACA TTGWCTCTAG 240  
 TGGATTTTCT TCCAAAATGA AAGTTGTTGC TTCTAGACTT TYAAGMKMRA TWKCCCMK 300  
 35 YWAWCKGAAC AMAMKCTGSW CYTCCWSYGC SKTRRMKRYC GYKSTATRRC WARWKSAYM 360  
 CCYGKMTGS RRGTAWYTSK TGCAYKAGG AACAAITGAG GAAGTTTGTT CTTTTTTCCA 420  
 40 TCGATACCA CAACTGCTTT TAGAACTTGA CAACGTAATT TCTGTTCTTT TTCAGAACAG 480  
 TAAAGAAAGG GGTAAGAAC TGAAGGAAAT CTGCCATTCT CAGTGGACAG GCAGGCATGA 540  
 TGCTTTTGAA ATTTTAGTGG AACTCCTGCA AGCACTTGTT TTAATGTTAG ATGGTATAAA 600  
 45 TAGTGACACA AATATTAGAT GGAATAACTA TATAGCTGGC CGAGCATTTG TACTCTGAGT 660  
 GCAGTGTGAG ATTTTGATTT CATGTTACT ATGTTGTTT TTAATAATGT CCTATCTTTT 720  
 50 ACAAGAGCCT TTGGGAAAAA CYCMAGGGG CAAACCTCTG ATGCTTCTTT TGCKKMSRT 780  
 ARMTTTTGAY ATRMARYACT RMTKSAITY AAYGRWGTGA CWSGAWAATA TTRAASTYTA 840  
 TACAATKAAT YWTRRYTSM KRMAGMYAAT CCGAAAYTGT GGMAAMYAAA CTTGATATTC 900  
 55 AAATGAAACT CCCTGGGAAA TTCCGCAGAG CTCACCAGGG TAACTTGGAA TCTCAGCTAA 960  
 CCTCTGAGAG TTAATAATAA GAAACCTTAA GTGTCCCAAC AGTGAGCAC ATTATTCAGG 1020  
 60 AACTTAAAGA TATATTCTCA GAACAGCACC TCAAAGCTCT TAAATGCTTA TCTCTGGTAC 1080

	CCTCAGTCAT GGGACAAC TC AAATTC AATA CGTCGGAGGA ACACCATGCT GACATGTATA	1140
5	GAAGTGACTT ACCCAATCCT GACACGCTGT CAGCTGAGCT TCATTGTTGG AGAATCAAAT	1200
	GGAAACACAG GGGGAAAGAT ATAGAGCTTC CGTCACCAT CTATGAAGCC CTCCACCTGC	1260
	CTGACATCAA GTTTTTCCT AATGTGTATG CATTGCTGAA GGTCTGTGT ATTCTTCCTG	1320
10	TGATGAAGGT TGAGAATGAG CGGTATGAAA ATGGACGAAA GCGTCTTAAA GCATATTGTA	1380
	GGAACACTTT GACAGACCAA AGGTCAAGTA ACTTGGCTTT GCTTAACATA AATTTTGATA	1440
15	TAAACACGA CCTGGATTTA ATGGTGGACA CATATATTAA ACTCTATACR AKTAMGTCAG	1500
	MGCTYYCTAC AKAYRAYTCM SWAMTGTGG AAARYWSSTA MGMSWGCWK TMMRRTMCG	1560
	GMWTFYYMK RKTGYAYMYW YGCGWMCAG AAAAAGCCGT AAGGTGTATG TAGACCACTT	1620
20	AATCACTAAA TATCTTTGCC TATAGGACTC CATTGAATAC ATTAGCCATT GATAATCTAC	1680
	CTGTTTAAAT GGGCCCTGTT TGAACCTCA AGCTTTGAAG ACCTACCTGT TCTTCCAGAA	1740
25	GAGAACGTTG AAAGTGCCAT GTTTCCTTTT GCGTGATCTC TGTGTATGGC ACTCTGGAAT	1800
	TGTTTCCAGT TTAAKTCATT TTAGACATAG CATTATTAT CACTGTGGAT CTCTACTTGT	1860
	TGGGTGTTAT GAATTCCTTG AAGAATATAT TTTGAAGAGG TGTGGGAGGA AGGAATACAT	1920
30	TTTATAAAAT GTTGTAGTGA AGCCACAAAT TGACCTTKGA CTAATAGGAG TTTTAAGTAT	1980
	GTAAAAATC TATACTGGAC AGTTACAAGA AATTACCGGA GAAAAGCTTG TGAGCTCACC	2040
35	AAACAAGGAT TTCAGTGTAG ATTTTGTCTT TCTTGAAC TT AAAGAAACAA ATGACAAAGT	2100
	TTGAATGGAA AAGCCTGCTG TTGTTCCACA TCTCGTTGCT GTTTACATTC CTTTGTGGAG	2160
	CCTACATCTT CCTAAGCTTT TTAGCAGGTA TATGTTGAAC ACTTCTGTTT CATGGTTGAG	2220
40	ACAGAATCAG AGGCCATGGA TACTGACAAC TGATTGTCT GTTTTTCCTT TCTGTCTTTT	2280
	TCCATGACTC TTATATACTG CCTCATCTTG ATTTATAAGC AAAACCTGGA AAACCTACAA	2340
45	AATAAGTGTT GTGGTTTATC TAGAAAAATA TGGAAAAATAT TGCTGTTATT TTTGGTGAAG	2400
	AAAATCAATT TTGTATAGTT TATTTCAATC TAAATAAAT GTGAATTTTG TTWATTAAA	2460
	AATTWGSAC AAABTBGHGG GGGDTCCAAA CHTWVTCGHG KAAMTCTCT WAARMATYTK	2520
50	ATAAACMSCT TCACAATTC	2539

55 (2) INFORMATION FOR SEQ ID NO: 110:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1751 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

60

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 110:

5	AGCATGAAGC CGATGGCCGT GGTGGCCAGT ACCGTCCTGG GCGTGGTGCA AAACATGCGT	60
	GCGTTTGGCG GGATCCTGGT GGIGGTCTAC TACGTATTTG CCATCATTGG GATCAACTTG	120
10	TTTAGAGGCG TCATTGTGGC TCTTCCTGGA AACAGCAGCC TGGCCCCGTC CAATGGGCTCG	180
	GCGCCCTGTG GGAGCTTCGA GCAGCTGGAG TACTGGGCCA ACAACTTCGA TGACTTTGCG	240
	GCTGCCCTGG TCACTCTGTG GAACTTGATG GTGGTGAACA ACTGGCAGGT GTTCTCTGGAT	300
15	GCATATCGGC GCTACTCAGG CCCGTGGTCC AAGATCTATT TTGTATTGTG GTGGCTGGTG	360
	TGCTCTGTCA TCTGGGTCAA CCTGTTCTG GCCCTGATTC TGGAGAAGTT CCTTCACAAG	420
20	TGGGACCCCC GCAGCCACCT GCAGCCCCCT GCTGGGACCC CAGAGGCCAC CTACCAGATG	480
	ACTGTGGAGC TCCTGTTTCA GGATATTCTG GAGGAGCCCG GGGAGGATGA GCTCACAGAG	540
	AGGCTGAGCC AGCACCAGCA CCTGTGGCTG TGCAGGTGAC GTCCGGGCTG CCATCCCAGC	600
25	AGGGGCGGCA GGAGAGAGAG GCTGGCCTAA CACAGGTGCC CATCATGGAA GAGGCGGCCA	660
	TGCTGTGGCC AGCCAGGCGA GAAGAGACCT TTCTCTGAC GGACCACTAA GCTGGGGACA	720
30	GGAACCAAGT CCTTTGCGTG TGGCCCAACA ACCATCTACA GAACAGCTGC TGGTGCTTCA	780
	GGGAGGCGCC GTGCCCTCCG CTTTCTTTTA TAGCTGCTTC AGTGAGAATT CCCTCGTCGA	840
	CTCCACAGGG ACCTTTTCTA CAAAAATGCA AGAAGCAGCG GCCTCCCTG TCCCTGTCAG	900
35	CTTCGGTGGT GCGTTTGCTG CCGGCAGCCC TTGGGGACCA CAGGCCTGAC CAGGGCCTGC	960
	ACAGGTTAAC CGTGAGTCTG TCTCATCTAT TCACAGCTGG GAATGATACT AATACCTCCG	1020
40	ATTTTAGCCC AGCACCACAG GGTACGTTCC AGTTTCTCTC TCTTTCCATA GCTGTAAAGC	1080
	CCTTCTGGG AATGGTCTC ATTCTCCTTA ATCTATTATT GGGTCAGTTT TCCTGCATGT	1140
	CCCCAGCCTC CCATCACTGC CACCCACTCC CCACAGAGAT GCCCTGCTCA TCCGACTGGG	1200
45	GCTTTGACTC CCACACTGTG TACCCCTCTT GTGTGGACGC CCGTGTGCCA AAACCTTCAG	1260
	CAAACAGCTT TCCAAATGGA AGTGTCACT GTCAGGCCTT TACAATCAGC AACAGCAAAA	1320
50	TCTACATGCT GCTGAGGGTC CTGCTCATT AAGATGCAAT AAATATGTAA GTACATAAAA	1380
	ACAGCAATAG AAGAAACGTA ATGCTTTATT CTCAAATATG ATGTCTACAT AGAAAAGCCA	1440
	AAATTATTAA GAATAGTAAG AATTCACCCA GCACTTTGGG AGGCCGAGGC GGGTGGATCA	1500
55	TGAGGTGAGG AGATCGAGAC CATCCTGGCT AACAGGGTGA AACCCGCTCT CTAATAAAAA	1560
	TACAAAAAAT TGGCCGGGCG CAGTGGCGGG CGCCTGTGGT CCCAGCTACT GGGGAGGCTG	1620
60	AGGCAGGAGA ATGGCGTGAA CCCGGGAAGC GGAGCTTGCA GTGAGCCGAG ATTGCGCCAC	1680

TGCAGTCCGC AGTCCAGCCT GGGCGACAGA GCGAGACTCC GTCTCAAAAA AAAAAAAAAA 1740  
 AAAAAAAAAA A 1751

5

(2) INFORMATION FOR SEQ ID NO: 111:

10

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1117 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 111:

AATGTTGTGG TGGTAGCATT TGGGTTAATT CTRATTATAG AGTCTCTTGG AGAGCAATGT 60  
 20 CCATAAACTA ATCCCAAACA ACATTGTCTT TTTRATGTTG TAGTGAACAG CAGAGAATTT 120  
 CAAAGGACCT TGCTAATATC TGTAAGACGG CAGCTACAGC AGGCATCATT GGCTGGGTGT 180  
 25 ATGGGGGAAT ACCAGCTTTT ATTCTATGCTA AACAACAATA CATTGAGCAG AGCCAGGCAG 240  
 AAATTATCA TAACCGGTTT GATGCTGTGC AATCTGCACA TCGTCTGCC ACACGAGGCT 300  
 TCATTCGTTA TGGCTGGCGC TGGGGTTGGA GAACTGCAGT GTTGTGACT ATATTCAACA 360  
 30 CAGTGAACAC TAGTCTGAAT GTATACCGAA ATAAAGATGC CTTAAGCCAT TTTGTAATG 420  
 CAGGAGCTGT CACGGGAAGT CTTTTTAGGA TAAACGTAGG CCTGCGTGGC CTGGTGGCTG 480  
 GTGGCATAAT TGGAGCCTTG CTGGGCACTC CTGTAGGAGG CCTGCTGATG GCATTTTCAGA 540  
 35 AGTACTCTGG TGAGACTGTT CAGGAAAGAA AACAGAAGGA TCGAAAGGCA CTCCATGAGC 600  
 TAAAACTGGA AGAGTGAAA GGCAGACTAC AAGTTACTGA GCACCTCCCT GAGAAAATG 660  
 40 AAAGTAGTTT ACAGGAAGAT GAACCTGAGA ATGATGCTAA GAAAATGAA GCACTGCTAA 720  
 ACCTTCCTAG AAACCTTCA GTAATAGATA AACAAGACAA GGACTGAAAG TGCTCTGAAC 780  
 TTGAAACTCA CTGGAGAGCT GAAGGGAGCT GCCATGTCCG ATGAATGCCA ACAGACAGGC 840  
 45 CACTCTTTGG TCAGCTGCT GACAAATTTA AGTGCTGGTA CCGTGGTGG CAGTGGCTTG 900  
 CTCTGTCTT TTTCTTTTCT TTTTAACTAA GAATGGGGCT GTGTACTCT CACTTTACTT 960  
 50 ATCCTTAAAT TTAATACAT ACTTATGTTT GTATTAATCT ATCAATATAT GCATACATGA 1020  
 ATATATCCAC CCACCTAGAT TTTAAGCAGT AAATAAACA TTTCGCAAAA GATTAAAGTT 1080  
 55 GAATTTTACA GTTAAAAAAA AAAAAAAAAA AAAAAA 1117

60

(2) INFORMATION FOR SEQ ID NO: 112:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1313 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 112:

10 GGCAGAGGTT TTCTTATATT TTAAGTAAAT TTAAAGTGGC TATCAGAATA TTTATTCTTG 60  
TTTGAGACTA CCAACATAAC TACGTGTTGA AGGTGCTTCA CAGAGAATAT ATTGCCTTTA 120  
ATGTGAAATA ATTTTCACCA ATGTTGCTAA CTTTAATAAA GTATAAAATT TGTAGAATAT 180  
15 TCAGTTAAGT AGTTGGTAAC CCTTTTCTAT TTTAGTAAAA CTTAATGCAT GTTTACTTTT 240  
TTTTGAAAGA TGCAGACAAT CTCTTTGAAC ATGAATTGGG GGCTCTCAAT ATGGCTGCAT 300  
TACTACGAAA AGAAGAAAGA GCAAGTCTTC TTAGTAATCT TGGCCCATGT TGTAAGGCGT 360  
20 TGTGCTTCAG ACGGGATTCT GCAATTCGAA AGCAGCTTGT TAAAAATGAG AAGGGCACCA 420  
TAAAACAAGC TTACACGAGT GCTCCAATGG TAGACAATGA ATTACTTCGA TTGAGTCTTC 480  
25 GGTATTTTAA GCGGAAGACT ACTTGCCATG CTCCAGGACA TGAAAAGACT GAAGATAATA 540  
AACTTTCACA GTCCAGTATC CAACAGGAAC TGTGTGTGTC TTAAGACCGA AGTTACAATA 600  
TGGTATTTTT GGTACTGTCT TCCTTCAGCA GTGCATATTC TTTTGCAAAG TTCTTTGGTT 660  
30 TGACAAGCAT TAGTGACAAA GGCAGAAAAG ATTTATCAGC CATGCTAAAA GAGTGAAGAA 720  
TTTTGATCTT TAGAGACACT AGTTTGGGCC AACTTAAGAT TTTACGTTAA TTTTACATA 780  
35 GTATTGACA CTCATGCAAA ATAATGTGAA AACATCTAGA TTTAGTAGTT TATTCTGCGC 840  
CTTTGTGTTA AACTGAAGAT TTTGGAAAAT GGTGTGCACT GCTCTCCAG CCTATGAATA 900  
TTTTTGTGAA ATGGAACCAT GGATTTATGT CTGGATCATC CATACAGAAC CAACAATTTT 960  
40 ATTCAAAAAC AATGTGTICA TCAAAGTAAT TGCTCACATT GTGCAGTACT ATGTTGTACA 1020  
GACCACGTGA AAGGGAATGC TGGTCTAGCT GCGTGGTAT GTTTATAGGC GAATTCAGC 1080  
45 AGAAGGAAGC CAAAATAGTT TTTTCTTTT GAAAGTTTTT TAAAAATTAT TTCATGGGTC 1140  
TTTTTTTTAA TTAATATGTG TGCATTGTTA CAATGTATGT TGGGATGTCT TTTGACCCTA 1200  
AATGCTTTTT TTGTTATCAG AGATTGTGTA CTATTTTAT TTTTAATAAA TGTATCTTCC 1260  
50 CTTTMAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAA 1313

55

## (2) INFORMATION FOR SEQ ID NO: 113:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1654 base pairs  
(B) TYPE: nucleic acid

60

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 113:

5	ACAGGGACAG AATACTTTCT TTCCTTCCTT CAAGTACAAG AAGGCTTTCT CTACCATTTG	60
	CGTCTACACT TTATTTTAAA AGCTATCCTT TTCTAGTAGT ATTTTATCAT GGCAATGGCA	120
10	TGATGACAAC AACAGTCTTT CATTACAGAC TGAAGGGAAG CATGTCCTTA CTAAAATAG	180
	TTCTGCTACT TTCCCTCCTA TTATAAGGAA ATTTTACAGA TTCTAAAAAT ACCTTAATT	240
15	TTCTTTGATT TTTATTTTAC CAAGTCACAA ATGTCCTTTT GATGTTTGA GAATTGTTCT	300
	CATAGAATCA CAAATCTGA CATTTTCATTA GATGATTATT TTCCTAGAAT CCCCAAAGAG	360
	CAGTGGCAGT CCATGGCTTG GTTGAAGCTA GAAATTTTCC TGCCCTGGT GACCTGGTAA	420
20	GCCTCCTGCT CGGAACCGTG TGAGTGGGTG AGGAAGATGA GAGATGGTCA GATGGAAGAG	480
	AGRAATACAT GAACTGCTCT GGCCTCTCTG GTTCTGTTCT TGGCCAGAG TTTTGA AAA	540
25	GCAGCGGANA TNGACTGACT TCACATGCTC AGCTTTCTCA GCCTTTTGTT TATTTGTG	600
	TCCTTAGATT TCCCTGTTGT AAAAGGGGCA AGAAAAGTAA CTCATCATCT CTAACACACC	660
	ATGGCAGCTT AGCCAGGTAG TCTTAGTGGT GGTGTTTAGG CATAAGATAT GCTGATCATC	720
30	AGTCTCAGGC CACAGTTTCC TTCACTAATC GTCCAGCTTG AGTGTTCTGT TCTCTCCTG	780
	CCCATTTCTT TGAACCTCCT GCTCTAGCCT TGGCGGAGGG AGAGTGCTAT TTGCTTTTGT	840
35	TTCTCCCTCTG TCTTAGGAAA AGCCATCTTT AATATAGTTC TTCACCACTG TTGGGGTTGT	900
	TTTGTGATTT TTTTCTCTT CCGAAGAACT CCTGGTTGTT ATTGGATTTT GTATTTTAAT	960
	ACAAATTATT GAATTTTATA AGCTTGTACA CAATATTTAA TTAGTGTGAA AGGAAACAAA	1020
40	GAATGCAGGA AAAATAATTT AATATCAACC TCAGTTGACA AGGTGCTCAG ATTATTCAAT	1080
	TGGGATCCT CCTTTTGTTA GGTTTTGTAG ACAACCCTAG ACCTAAACTG TGTCACAGAC	1140
45	TTCTGAATGT TTAGGCAGTG CTAGTAATTT CCTCGTAATG ATTCTGTTAT TACTTCTTA	1200
	TTCTTTATTC CTCTTCTTC TGAAGATTAA TGAAGTTGAA AATTGAGGTG GATAAATACA	1260
	AAAAGGTAGT GTGATAGTAT AAGTATCTAA GTGCAGATGA AAGTGTTTA TATACATCCA	1320
50	TTCAAAATTA TGCAAGTTAG TAATTACTCA GGGTTAACTA AATTACTTTA ATATGCTGTT	1380
	GAAYCTACTC TGTCCTTGG CTAGAAAAAA TTATAACAG GACTTTGTAG TTTGGGAAGC	1440
55	CAAAITGATA ATATTCTATG TTCTAAAAGT TGGGCTATAC ATAAATTATT AAGAAATATG	1500
	GATTTTATT CCCAGGATAT GGTGTTCAAT TTATGATATT ACGCAGGATG ATGTATTGAG	1560
	TAAAATCAGT TTTGTAAATA TGTAATATG TCATAAATAA ACAATGCTTT GACTTATTT	1620
60	CAAAAAAAA AAAAAATAAA NTTCGAGGGG GGGC	1654

## 5 (2) INFORMATION FOR SEQ ID NO: 114:

## (i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 1171 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 114:

15 GGCAAACTTT CCCCAANGC TTCGAACTT GCAAGCCGAA ACCTTGAATC GTTAAAAGTT 60  
GGGTTGCGNC GCGCCCTGG CCCGAAGAAG CGCAATTGGC GTTCCGCGAA CGTTGGCCCT 120  
CAACGGCTCG GCAGCCAGCC ATGTCTGCA CCCAGGACAG CGGCCCTGGG CTACAAGGAC 180  
20 CTGGMCTCA TCTTCTGCG CCGACCTGCG CGGGTAAGG GGWAGTTTCA GACTGTGAAG 240  
GACGTCTGTC TGGACTGCCT GTTGGACTTC TTACCCGAGG GGGTGAACAA AGAGAAGATC 300  
25 ACACCACTCA CGCTCAAGGA AGCTTATGTG CAGAAAATGG TTAAAGTGTG CAATGACTCT 360  
GACCGATGGA GTCTTATATC CCTGTCAAAC AACAGTGGCA AAAATGTGGA ACTGAAATTT 420  
GTGGATCCC TCCGAGGCA GTTGAATTC AGTGTAGATT CTTTCAAAT CAAATTAGAC 480  
30 TCTCTCTGC TCTTTTATGA ATGTTAGAG AACCCAATGA CTGAGACATT TCACCCACA 540  
ATAATCGGG AGAGCGTCTA TGGCGATTTC CAGGAAGCCT TTGATCACCT TTGTAACAAG 600  
35 ATCATTGCCA CCAGGAACCC AGAGGAAATC CGAGGGGGAG GCCTGCTTAA GTACTGCAAC 660  
CTCTTGGTGA GGGGCTTTAG GCGCCCTCT GATGAAATCA AGACCTTCA AAGGTATATG 720  
TGTTCCAGGT TTTTCATCGA CTTCTCAGAC ATTGGAGAGC AGCAGAGAAA ACTGGAGTCC 780  
40 TATTTGCAGA ACCACTTTGT GGAATTTGGA AGACCGCAAG TATGAGTATC TCATGACCCT 840  
TCATGGAGTG GTAAATGAGA GCACAGTGTG CCTGATGGGA CATGAAAGAA GACAGACTTT 900  
45 AAACCTTATC ACCATGCTGG CTATCCGGGT GTTAGCTGAC CAAAATGTCA TTCCTAATGT 960  
GGCTAATGTC ACTTGCTATT ACCAGCCAGC CCCCTATGTA GCAGATGCCA ACTTTAGCAA 1020  
TTACTACATT GCACAGGTC AGCCAGTATT CACGTGCCAG CAACAGACCT ACTCCACTTG 1080  
50 GCTACCCTGC AATTAAGAAT CATTTAAAAA TGTCTGTGG GGAAGCCATT TCAGACAAGA 1140  
CAGGAGAGAA AAAAAAAAAA AAAAAAAAAA A 1171

55

## (2) INFORMATION FOR SEQ ID NO: 115:

- 60 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 842 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 115:

GGTCTGCGCC GGAAGTGCAT GAGCTGCCGA TGTGGTGCCT AGTGATTGCG GTTTCGGTCG 60  
10 CTCTCCCGTG TTTCCCGGGC TGGGTATTG CCTCGCACCA TGGCGCCCAA GGGCAAAGTG 120  
GGCACGAGAG GGAAGAAGCA GATATTGAA GAGAACAGAG AGACTCTGAA GTTCTACCTG 180  
15 CGGATCATAC TGGGGGCCAA TGCCATTAC TGCCTTGTA CGTTGGTCTT CTTTACTCA 240  
TCTGCCCTCAT TTTGGGCTG GTTGGCCCTG GGCTTTAGTC TGGCAGTGTA TGGGGCCAGC 300  
TACCACTCTA TGAGCTCGAT GGCACGAGCA GCGTCTCTG AGGATGGGGC CCTGATGGAT 360  
20 GGTGGCATGG ACCTCAACAT GGAGCAGGGC ATGGCAGAGC ACCTTAAGGA TGTGATCCTA 420  
CTGACAGCCA TCGTCAGGT GCTCAGCTGC TTCTCTCTCT ATGTCCTGGTC CTTCTGGCTT 480  
25 CTGGCTCCAG GCGGGGCCCT TTACCTCCTG TGGGTGAATG TGCTGGGCC CTGGTCACT 540  
GCAGACAGTG GCACCCAGC ACCAGAGCAC AATGAGAAAC GGCAGCGCG ACAGGAGCGG 600  
CGGCAGATGA AGCGTTATA GCCATTGACA TTGTGGCCAC AGGCCACTGG CCCTGGGTGG 660  
30 CTCTGTCAGG GTGCACAGCC CCTCATGCCT GGAGCAATGA GGGTCTAGTC CAGGGGCCAA 720  
AAGCAGTCTG AGGTATTGGG TATACTTATA CTCTATAGGG TCGTTGAATA AATGGCTTAG 780  
35 AATGTGAAAA AAAAAAAAAA AAAAACTCG AGGGGGGCC GGTACCCAAT TTCNCCTANA 840  
AT 842

40

(2) INFORMATION FOR SEQ ID NO: 116:

(i) SEQUENCE CHARACTERISTICS:

- 45 (A) LENGTH: 1640 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 116:

GGCACGAGGC GCGGCGAGCG GTGGCGGCGG CGCCCCCGG CCGGAGCCGT TCCCTTTCCC 60  
GTGCGGGAGC GCGGGGYCGG GGGCCAGGG ACCCGGGCC ACGGAGAGCG GGAAGAGGAT 120  
55 GGATTGCCCC GCCTCCCCC CCGGATGGAA GAAGGAGGAA GTGATCCGAA AATCTGGGCT 180  
AAGTGCTGGC AAGAGCGATG TCTACTACTT CAGTCCAAGT GGTAGAAGT TCAGAAGCAA 240  
GCCTCAGTTG GCAAGGTACC TGGGAAATAC TGTGATCTC AGCAGTTTTC ACTTCAGAAC 300  
60



	TGGAAAGATG ATGCCTAGTA AATTACAGAA GAACAAACAG AGACTGCGAA ACGATCCTCT	360
	CAATCAAAT AAGGGTAAAC CAGACTTGAA ATACAACATT GCCAATTAGA CAAACAGCAT	420
5	CAATTTTCAA ACAACCGTA ACCCAAAGTC ACAAATCATC CTAGTAATAA AGTGAAATCA	480
	GACCCACAAC GAATGAATGA ACAGCCACGT CAGCTTTTCT GGGAGAAGAG GCTACAAGGA	540
10	CTTTAGTGCA TCAGATGTAA CAGAACAAAT TATAAAAACC ATGGAACCTAC CCAAAGGTCT	600
	TCAAGGAGTT GGTCCAGTAG CAATGATGAG ACCCTTTTAT CTGCTGTTGC CAGTGTCTTG	660
	CACACAAGCT CTGCGCCAAT CACAGGGCAA GTCTCCGCTG CTGTGGAAAA GAACCTGCTG	720
15	TTTGGCTTAA CACATCTCAA CCCCTCTGCA AAGCTTTTAT TGTACAGAT GAAGACTCAG	780
	GAAACAGAAG AGCGAGTACA GCAAGTACGC AAGAAATTGG AAGAAGCACT GATGGCAGAC	840
20	ATCTTGTCGC GAGCTGCTGA TACAGAAGAG ATGGATATTG AAATGGACAG TGGAGATGAA	900
	GCCTAAGAAT ATGATCAGGT AACTTTCGAC CGACTTTCCC CAAGAGAAAA TTCCTAGGAA	960
	ATTGAACAAA AATGTTTCCA CTGGCTTTTG CCTGTAAGAA AAAAAATGTA CCCGAGCACA	1020
25	TAGAGCTTTT TAATAGCACT AACCAATGCC TTTTATAGTG TATTTTGTAT GTATATATCT	1080
	ATTATTCAAA AAATCATGTT TATTTTGAGT CCTAGGACTT AAAATTAGTC TTTTGTAAATA	1140
30	TCAAGCAGGA CCCTAAGATG AAGCTGAGCT TTTGATGCCA GGTGCAATCT ACTGGAAATG	1200
	TAGCACTTAC GTAAAACATT TGTTTCCCCC ACAGTTTAA TAAGAACAGA TCAGGAATTC	1260
	TAAATAAATT TCCAGTTAA AGATTATTGT GACTTCACTG TATATAACA TATTTTATA	1320
35	CTTTATTGAA AGGGGACACC TGTACATTCT TCCATCGTCA CTGTAAAGAC AAATAAATGA	1380
	TTATATTCCA CAGAAAAAA AAAAAAAW MWSTYGARRR GSRGCMCRSW AYMMARWCC	1440
40	CCWMRIWRGS MKTCSIMKA YTTACATTCA ACTCTGATCC CGGGCCCTTA GGTTTGACAT	1500
	GGGAGGTGGG AGGAAGATAG CGCATATATT TGCAGTATGA ACTATTGCCT CTGGGACGTT	1560
	GTGAGGAATT GTGCTTTCAC CAGAATTTCT AAGGATTTCT GGCTTAAATA TCACCTAGCC	1620
45	TGTGGTAATT TTTTTCCT	1640

50 (2) INFORMATION FOR SEQ ID NO: 117:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 952 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 117:

60 TGAATTTAGN AAACACTTTG GAAACTCAT AACCTCATCA GAAACTGCCT TTAGCCACAC 60

	TCCTGACCTT CTAGATGAGT AACAAAAAA TGAAATAAGT TCTTGGAAT TAAGCCATTT	120
5	ATTTTAATTT GCTATTTTTT TCAATGTTCT AGGTATCTTT AAATTTGTTA TTGTGGAATC	180
	ATTTTCCTGC CAGATACCTT TATCAAAATT ATGGCCTCA TGAGAGCTGA AGTAAGTCAG	240
	CTTTTGGTG AACTTTAGTG GACTTCTGTG AGATTGTAGT TGTACTTTGT ATCTCTAAAT	300
10	CTAAAGATAG TTTTTTAAAA CTCCTAAAGA AAATCTGCTC TCCTTTCTGA TCTAAAAACT	360
	CATCTTTGGG GTAAAGAGTT AAGTGTCAA AGGTTGTCAC AGTTCATGAG GTCAGAGGGA	420
15	GCTAGCCTGG CACCTGGACT CTGCCCATCC ACAGCTGACA GATTCCAACA GAAGTGATTT	480
	TAAATCTCC AGTAGACAAT GCTGGTAAG GGAGGGGTA GGGCTGGGTT ATTAAGATAC	540
	AGGCTGCTGT ATTTTACATT GGTGTGGGG GAAGGGGAGC CTGGAGAAA CAAAGTCACT	600
20	ATTCCTTTT TTGAAACAGG AAAAAAATT ATTTTGTGT CAGTAAAAAT GGTAGAGAAT	660
	TCCAATGTCC CTAGCCACAA GGGACCAGTT CCACTGAGAA GTGAACAGTG GGAACCAAA	720
25	ATTTCAGAAA CATTGGGGGA AGGAAAATT GGCTTTCTCT TAATTGGCAG ATGTTCCAGT	780
	GGGGSGGGG GGCTCTGTTT TTGTTGGGAT GTGTTATGTT GTATGTACGC ATATATGGAC	840
	CGGAGTCTGC TGAGTTTATA AGGTTCCAAA AATATGGTAA AATCTTGGTT TTTGTTAATT	900
30	TATCTCAATA AAAGCCCACT GGRACCTCAA AAAAAAAAA AAAAAAAGA NN	952

35 (2) INFORMATION FOR SEQ ID NO: 118:

(i) SEQUENCE CHARACTERISTICS:

- 40 (A) LENGTH: 1256 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 118:

45	GACGTCATAG GTAAACAGGC TCTGTATCCG TGGCAGCGGC CGTGGCAGGC TGGCTGGGTA	60
	CCGGCTGTGC CTGACCCAGG AGAAGCTGCC TGTCTACATC AGCCTGGGCT GCAGCGCGCT	120
50	GCCGCCGCGG GCGCGGCAGC TGAATATGT GCTCTTCAGG GCGGGCACCG TGTTGCATTTC	180
	ATCTTTGTAC CCCCAGCATC TAGCAGTGTG GGCATGTAGT AGGCACTCAA GAAATGTGTG	240
	TTGAATGAAC GATGCTGTG ACAAGCAAGC GGACTTTATT CTTCCTGAC CCTTGCTCCT	300
55	ATGACACACC TCCTCCTGAC TGCCACTGTC ACTCCTTCAG AGCAGAACTC CTCTAGGGAA	360
	CCTGGATGGG AAACAGCCAT GCCAAGGAC ATCCTGGGTG AAGCAGGGCT ACACTTTGAT	420
60	GAAGTGAACA AGCTGAGGGT GTTGACCCA GAGGTTACCC AGCAGACCAT AGAGCTGAAG	480

	GAAGAGTGCA AAGACTTTGT GGACAAAATT GCCAGTTTC AGAAAATAGT TGGTGGTTTA	540
	ATTGAGCTTG TTGATCAACT TGCAAAAGAA GCAGAAAATG AAAAGATGAA GGCCATCGGT	600
5	GCTCGGAACT TGCTCAAATC TATAGCAAAG CAGAGAGAAG CTCAACAGCA GCAACTTCAA	660
	GCCCTAATAG CAGAAAAGAA AATGCAGCTA GAAAGGTATC GGGTTGAATA TGAAGCTTTG	720
10	TGTAAAGTAG AAGCAGAACA AAATGAATTT ATTGACCAAT TTATTTTTC A GAAATGAACT	780
	GAAAATTTCTG CTTTTATAGT AGGAAGGCAA AACAAAAAAA AGCCTCTCAA AACCAAAAAA	840
	ACCTCTGTAG CATTCCAGCG GCTTGACCAA TGACCTATGT CACAAGAGGT GGCCTGTAAG	900
15	GAATGCAGCC CCCTGAAGAC AGCACTACAA GTCTGGGGGA GCCAGTTTTA ACATCAGTGC	960
	ACAGCTGCTG CTGGTGGCCC TGCAGTGAC GTTCTCACCT CTTATGCTTA GTTGGAAC TA	1020
20	AGCAGTTTGT AAACTTTCAT CCTTTT TTTT GTAAATTCAC AAAGCTTTGG AAGGAGAAGC	1080
	AATAAATTTT TGT TTTTCAA TGGCTTGATG TACCTTTTCT CCTGTGCTC TTGAAATATG	1140
	TTTAACCTCT CATGAGAGAA CCCTGGATTG TCTATCCCTT AGTCCACAAA ACAACCAGG	1200
25	CAGTGGTCAG CAGCTACCTT TNATTGGAT CACACACGTG AGTCAGACAG TACCAC	1256

30 (2) INFORMATION FOR SEQ ID NO: 119:

(i) SEQUENCE CHARACTERISTICS:

- 35 (A) LENGTH: 1143 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 119:

40	GGCCGTAGCA GCCGGGCTGG TCCTGCTGCG AGCCGGCGGC CCGGAGTGGG GCGGCGCAT	60
	GTACCTTCCA CATTGAGTAT TCAGAAAGAA GTGATCTGAA CTCTGACCAT TCTTTATGGA	120
45	TACATTAAGT CAAATATAAG AGTCTGACTA CTTGACACAC TGGCTCGAGC AAACATGAAC	180
	GTTGGAGTTG CCCACAGTGA AGTGAATCCA AATACCCGTG TCATGAACAG CCGGGGTATG	240
	TGGCTGACAT ATGCATTGGG AGTTGGCTTG CTTTATATG TCTTACTCAG CATTCCCTTC	300
50	TTTCTGTTTC CTGTTGCTTG GACTTTAACA AATATTATAC ATAATCTGGG GATGTACGTA	360
	TTTTTGATG CAGTGAAAGG AACACCTTTC GAACTCCTG ACCAGGGTAA AGCAAGGCTC	420
55	CTAACTCATT GGAACAACCT GGACTATGGA GTACAGTTTA CATCTTCAG GAAGTTTTC	480
	ACAATTCTC CAATAATTCT ATATTTCTG GCAAGTTTCT ATACGAAGTA TGATCCAAC T	540
	CACTTCATCC TAAACACAGC TTCTCTCCTG AGTGTACTAA TTCCCAAAAT GCCACAAC TA	600
60	CATGGTGTTC GGATCTTTGG AATTAATAAG TATTGAAATG TTTTGAACT GAAAAAAAT	660

	TTTACAGCTA CTGAATTTCT TATAAGGAAG GAGTGGTTAG TAAACTGCAC TGTTCCTSTG	720
	ATAATGTGAA ATGAGAAGTA TTTACATTGG AGGGCCAATG GCTGGTCCTT CAAGTGCTGT	780
5	TTTGAAGTGC AGATTTCAT TAAATGATGC CTCTGTTTAA TACACCTGGT ACATTTCTGA	840
	AGAGGGGCTT TATAAGCAGG CTGGGCAGGC CCAGCTTATA AGTTAAAGGG CATCACAGTG	900
10	AGGGTGTAGT AGATAAATTC AAGGAAATAA GAGATTGTGA AGAACTAGG ACCAGCTTAA	960
	CTTATAATGA ATGGGCATG TGTTAAGAAA AGAACATTTC CAGTCATTCA GCTGTGGTTA	1020
15	TTTAAAGCAG ACTTACATGT AAACCGAAT CCTCTCTATA CAAGTTTATT AAAGATTATT	1080
	TTTATTACCG TAAAAAANA AAAAAAANA AAAAAAANA AAAAAAANA AAAAAAANA	1140
	GAN	1143

20

## (2) INFORMATION FOR SEQ ID NO: 120:

25

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1782 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

30

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 120:

	CAGGCCCCCG CCCCCACCC ACCTCTGCGT TGCTGCCCCG CCTGGGCCRG GCCCCAAAGG	60
35	CAAGGACAAA GCAGCTGTCA GGGAACTCC GCCGGAGTCG AATTACGTG CAGCTGCCGG	120
	CAACCACAGG TTCCAAGATG GTTTCGGGG GCTTCGGCTG TTCCAAGAAC TGCCTGTGCG	180
40	CCCTCAACCT GCTTTACACC TTGGTTAGTC TGCTGCTAAT TGAATTGCT GCGTGGGGCA	240
	TTGGCTTCGG GCTGATTCC AGTCTCCGAG TGGTCGGCGT GGTCAATTGCA GTGGGCATCT	300
	TCTTGTTCTT GATTGCTTTA GTGGGTCTGA TTGGAGCTGT AAAACATCAT CAGGTGTTGC	360
45	TATTTTTTTA TATGATTATT CTGTTACTTG TATTTATTGT TCAGTTTCT GTATCTTCCG	420
	CTTGTTTAGC CCTGAACCAG GAGCAACAGG GTCAGCTTCT GGAGGTGGT TGAACAATA	480
50	CGGCAAGTGC TCGAAATGAC ATCCAGAGAA ATCTAACTG CTGTGGGTC CGAAGTGTTA	540
	ACCCAAATGA CACCTGTCTG GCTAGCTGTG TTAAAAGTGA CCACTCGTGC TCGCCATGTG	600
	CTCCAATCAT AGGAGAATAT GCTGGAGAGG TTTTGAGATT TGTGGTGGC ATTGGCCTGT	660
55	TCTTCAGTTT TACAGAGATC CTGGGTGTTT GGCTGACCTA CAGATACAGG AACCAGAAAG	720
	ACCCCGCGC RAATCCTAGT GCATTCTTTT GATGAGAAAA CAAGGAAGAT TTCCTTTCGT	780
60	ATTATGATCT TGTTCACCTT CTGTAATTTT CTGTTAAGCT CCATTTGCCA GTTTAAGGAA	840

GGAAACACTA TCTGGAAAAG TACCTTATTG ATAGTGAAT TATATATTTT TACTCTATGT 900  
 TTCTCTACAT GTTTTTTCT TCCGTGCT GAAAAATATT TGAACTTGT GGTCTCTGAA 960  
 5 GCTCGGTGGC ACCTGGGAAT TACTGTATT CATGTGCGG CACTGTCCAC TGTGGCCTTT 1020  
 CTTAGCATTT TTACCTGCAG AAAAAGTTG TATGGTACCA CTGTGTTGGT TATATGGTGA 1080  
 10 ATCTGAACGT ACATCTCACT GGTATAATTA TATGTAGCAC TGTGCTGTGT AGATAGTTCC 1140  
 TACTGGAAAA AGAGTGGRAA TTTATTAAAA TCAGAAAGTA TGAGATCCTG TTATGTTAAG 1200  
 GGAAATCCAA ATTCCCAATT TTTTGGTC TTTTAGGAA AGATGTGTG TGGTAAAAAG 1260  
 15 TGTTAGTATA AAAATGATAA TTWACTKGTA GTCMTTATG ATWACACCAA TGTATTCTAG 1320  
 AAATAGTTAT GYCYTAGGAA ATTGTGGTTT AATTTTTGAC TTTTACAGGT AAGTGCAAAG 1380  
 20 GAGAAGTGGT TTCATGAAAT GTTCTAATGT ATAATAACAT TTACCTTCAG CCTCCATCAG 1440  
 AATGGAACGA GTTTGTAGTA ATCAGGAAGT ATATCTATAT GATCTTGATA TTGTTTATA 1500  
 ATAATTGAA GTCTAAAAGA CTGCATTTT AAACAAGTA GTATTAATGC GTTGGCCAC 1560  
 25 GTAGCAAAAA GATATTTGAT TATCTTAAAA ATGTTAAAT ACCGTTTCA TGAAAGTTCT 1620  
 CAGTATTGTA ACAGCAACTT GTYAAACCTA AGCATATTG AATATGATCT CCCATAATTT 1680  
 30 GAAATTGAAA TCGTATTGTG TGGCTCTGTA TATCTGTGA AAAAATTAAA GGACAGAAAC 1740  
 CTTTCTTGT GTATGCATGT TTGAATTAAA AGAAAGTAAT GG 1782

35

(2) INFORMATION FOR SEQ ID NO: 121:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 610 base pairs  
 40 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 121:

45 GTTGGCTGCA GATTTGTGGT GCGTCTGAG CCGTCTGTCC TGCGCCAAGA TGCTTCAAAG 60  
 TATTATTAAA AACATATGGA TCCCATGAA GCCCTACTAC ACCAAAGTTT ACCAGGAGAT 120  
 50 TTGGATAGGA ATGGGGCTGA TGGGCTTCAT CGTTTATAAA ATCCGGGCTG CTGATAAAAG 180  
 AAGTAAGGCT TTGAAAGCTT CAGCGCCTGC TCCTGGTCAT CACAACCAGA TTTACTTGGA 240  
 55 GTACATGTGA AAGAAAACGT CAGTCTGCCT GTAAATTCA GCAAGCCGTG TTAGATGGGG 300  
 AGCGTGAAC GTCACGTGAC ACTTGTATAA GTACCGTTTA CTTTCATGGCA TGAATAAATG 360  
 GATCTGTGAG ATGCACTGCT ACCTGGTACT GCTTTCAGTG TGTCCCCCT CAGCCCTCCG 420  
 60 GCGTGTGAG CATACTCTGA GTAGATAATT TGTCATGCAG CGCATGCAAT CAGAACTCA 480

CTGAGCCACC CATCATTTGTG AAATAATTAC CTCAGTTGTA CAGGACTTGG TGATCAGGAT 540  
CCAGGCACTC ACTTGTATTTC TACTGCTCAA TAAACGTTTA TTAAACTTGA AAAAAAAAAA 600  
5 AAAAAAAAAA 610

10

(2) INFORMATION FOR SEQ ID NO: 122:

(i) SEQUENCE CHARACTERISTICS:  
15 (A) LENGTH: 526 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 122:

GGTACGCCCTG CAGGTACCGG TCCGGAATTC CGGGTCGCCC ACGCGTCNGG CCACGCGTCC 60  
ACCCACGCGT CCGSCCAGCG GTCGGAGCCG AGCCGGA CTG CAGGATGA TCACGGACGT 120  
25 GCAGCTCGCC ATCTTCGCCA ACATGCTGGG CGTGTGCTC TTCTTGCTTG TCGTTCCTA 180  
TCACTACGTG GCCGTCAACA ATCCCAAGAA GCAGGAATGA AAGTGGGCT TTCTCCGCC 240  
CAGGGTTCCA GGACATAGTC TGAGGCAAGA TGGAGGTAT GAGGGGCTT CACACTTCAC 300  
30 TTCATCCCTT CTACCCATCA CAACATACAA AGCAACTACA CCTGGATTTT TCCAAACAAC 360  
TTTTATTTCC TCAGAGTCTT CCTTAATCCT ATGGAACAAG AAGCTGCCAC TGAATAGGGC 420  
35 CCAGTATAGG GGCTTGCTTT TCTACTCCCT CCCCCAATA TAAAAATATA GACTTTTAA 480  
AAAAAAAAA AAAAANITCG NGGGGGGSCC GGTACCCATC CCCCTA 526

40

(2) INFORMATION FOR SEQ ID NO: 123:

(i) SEQUENCE CHARACTERISTICS:  
45 (A) LENGTH: 2081 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 123:

TGTACCGGTC CGGAAATTCC CGGGTCGACC CAGTCGTCS GGGGAACATG GCGGCTKCGG 60  
AGCCGGCGGT CCTTGCCTC CCCAACAGCG GCGCCGGGGG CGCGGGGGCG CCGTCGGGCA 120  
55 CAGTCCCGGT GCTCTTCTGT TTCTCAGTCT TCGCGGACC CTCGTGGTG CCACACGGGG 180  
CGGGCTACGA GCTGCTCATC CAGAAGTCC TCAGCCTGTA CGGCGACCAG ATCGACATGC 240  
60 ACCGCAAAAT CTGGTGCAG CTGTTCCCGG AGGAGTGGG CCAGTACGTG GACTTGCCCA 300

	AGGGCTTCCC GGTRAGCGAG CGCTGCAAGG TCGGCTCGT GCCGYTGCAG ATCCAGCTCA	360
5	CTACCCCTGGG AAATCTTACA CCTTCAAGCA CTGTGTTTTT CTGCTGTGAT ATGCAGGAAA	420
	GGTTCAGACC AGCCATCAAG TATTTTGGGG ATATTATTAG CGTGGGACAG AGATTGTTGC	480
	AAGGGGCCCG GATTTTAGGA ATTCTGTTA TTGTAACAGA ACAATACCCT AAAGGTCTTG	540
10	GGAGCACGGT TCAAGAAATT GATTTAACAG GTGTAAACT GGTACTTCCA AAGACCAAGT	600
	TTTCAATGGT ATTACCAGAA GTAGAAGCGG CATTAGCAGA GATTCCCGGA GTCAGGAGTG	660
15	TTGTATTATT TGGAGTAGAA ACTCATGTGT GCATCCAACA AACTGCCCTG GAGCTAGTTG	720
	GCCGAGGAGT CGAGGTTTAC ATTGTTGCTG ATGCCACCTC ATCAAGAAGC ATGATGGACA	780
	GGATGTTTGC CCTCGAGCGT CTCGCTCRAR CCGGGATCAT AGTGACCACG AGTGAGGCTG	840
20	TTCTGCTTCA GCTGGTAGCT GATAAGGACC ATCCAAAATT CAAGGAAATT CAGAATCTAA	900
	TTAAGGCGAG TGCTCCAGAG TCGGGTCTGC TTTCCAAAGT ATAGGACATT TGAAGAACTG	960
25	GTATGCTACT CACTGGTGAA GGACAGTCAG GTGAAGGACT GTAAGCCAC ACAAGCTCTT	1020
	CTTATCTCTA CTAGAATTAA AATGTTAAGT CAAAAACGGC TCCTTTTTTG CGCCTCCTAG	1080
	TGAAACTTAA CCAGCTAGAC CATTTGAGTA CCAGCATTTA GTTACAAACG TCAAAGGCTT	1140
30	CCGGTGCTGC TTACCTTCCT TTTTGTGTTA TGTGCTTTTA TTTATTAAAA AAAATTACAA	1200
	TGAAGATGCC TGTMTTGTCT CTACTGTGTA CTCTGATCGT ATCTTTCCAA AGTGCAGACT	1260
35	CTTGTGAAGT TTTCTTAAAT TGTTCACTTT AAAGAAAATG ACGTACCAAC AATGATTGG	1320
	CTTTTATATT ACTGTAAGAT GTTATAATGT TAATGTGGAT GTAGTGCTTT TACTTTACAG	1380
	ATTGATTGGA ATAAGATTAT TGCATATGAA TTTACCCACA GGACTCTGAA TCATGTTACC	1440
40	CACTCCCTC ACAATGTTGT CCACTTAGTG AGTTGCATTG ATCTATCCGT ACCAAATGAT	1500
	GTGAATAAT TACATATCTT TCTTGACTAT ACTGATTTCT TATTTTGGTC ACTATTACTA	1560
45	AATCTCTGTT AATATTCTCT CTTTAACTG AAAAGGGATG GGATAGAAGG GTTTGCAATG	1620
	CCATATTATT GGTGGAGGGC TGTTTTAACA TCTTTGAAGT ATGGCTTGCT GAATATCTTT	1680
	ACCAACATCT TGAATATATA TTCTAGTGT CACAAGATTT AGCAAAAAGA TAAAGCTTGG	1740
50	GTGGAATATC ATTTTAAAAT GTTCATGTTT TGTCTATAT TTTCTTCACC TACTCTCCAA	1800
	ATATTGTAAT GCAAAAAGTC TCAGTAATGA TTTGGTAGTA TTAATTTTGT GGTCATTGTT	1860
55	TCTCTCGAT AAATTATT TTTCATTAAATA CTTRITAGAG GGTTTTGAAA TGTTTTCAA	1920
	ATATGTGAAA TGTGAACTG CTGTCTTTTA TATTAAAGTA ATTAAAGAAA ATGTATTGTG	1980
	ATTGAAATTA TTTTGNCTC CACAAGATGG CTCTATGAGT ATTCTTCAG GGATTCTAAT	2040
60	ATTTATTTAA GGTNATAAAA TCTTGACATT TATAATCTTT C	2081

## 5 (2) INFORMATION FOR SEQ ID NO: 124:

## (i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 1717 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 124:

15 CCCC GGCGGA GCTGGACCCG CGGTGGGCTA GGGGCAGGGC CGGAGCCGCG GCGGCGGAGC 60  
TGTGGATCCT TCATGATGAG AGATTTGGGG ACACTTCTCT CTCTGTGTG TAGTTGATAG 120  
TTTGGTGGTG AAGAGATGGC TGACAGTGTG AAAACCTTTC TCCAGGACCT TGCCAGAGGA 180  
20 ATCAAAGACT CCATCTGGGG TATTTGTACC ATCTCAAAGC TAGATGCTCG AATCCAGCAA 240  
AAGAGAGAGG AGCAGCGTCG AAGAAGGGCA AGTAGTGTCT TGGCACAGAG AAGAGCCCAG 300  
25 AGTATAGAGC GGAAGCAAGA GAGTGAGCCA CGTATTGTGA GTAGAATTTT CCAGTGTGTG 360  
GCTTGAATG GTGGAGTGT CTGGTTCAGT CTCCTCTTGT TTTATCGAGT ATTTATTCCT 420  
GTGCTTCAGT CGTAACAGC CGAATTATC GGTGACCCAT CACTACATGG AGATGTTTGG 480  
30 TGTGGCTGG AATTCCTCCT CACGTCAATT TTCAGTGCTC TTTGGGTGCT CCCCTGTGTT 540  
GTGCTTAGCA AAGTGGTGAA TGCCATTTGG TTTCAGGATA TAGCTGACCT GGCATTTGAG 600  
35 GTATCAGGGA GGAAGCCTCA CCCATTCCTT AGTGTGAGCA AAATAATTGC TGACATGCTC 660  
TTCAACCTTT TGCTGCAGGC TCTTTTCCTC ATTCAGGGAA TGTTGTGAG TCTCTTTCCC 720  
ATCCATCTTG TCGGTGAGCT GGTAGTCTC CTGCATATGT CCCTTCTCTA CTCACTGTAC 780  
40 TGCTTTGAAT ATCGTTGGTT CAATAAAGGA ATTGAAATGC ACCAGCGGTT GTCTAACATA 840  
GAAAGGAATT GGCCTTACTA CTTTGGGTTT GGTGTGCCCT TGGCTTTTCT CACAGCAATG 900  
45 CAGTCTCAT ATATTATCAG TGGCTGCCCT TTCTCTATCC TCTTTCCTTT ATTCATTATC 960  
AGCGCCAATG AAGCAAAGAC CCCTGGCAAA GCRTATCTCT TCCAGTTGCG CCTCTTCTCC 1020  
TTGGTGGTCT TCTTAAGCAA CAGACTCTTC CACAAGACAG TCTACCTGCA GTCGGCCCTG 1080  
50 AGCAGCTCTA CTTCTGCAGA GAAGTCCCT TCACCGCATC CGTCGCCTGC CAAACTGAAG 1140  
GCTACTGCAG GTCAGTGTG TGCTGCCAT CCAAAGGGGA TGGCGGGGAT TGAAGAAGC 1200  
55 TGTGGCAGCT CTTTTCCTG TTCACCTCCC GCCTGCCAGG GAAGGCAGGA CCCGCTCTGC 1260  
CAAGGGCCCT CTGCGTATTC CCTTCTCTCT GAGGAATTGA AATTTTGTG TCTGGTGCAC 1320  
60 GTAAGGCAGA ATGTTCCCTG ACACCAGTGT GTGGATTTTT AACATCACCG TGAGTCTGAA 1380



AGGACCACAG GTTTTCTGC AGCTATTTTC TAGCATTTGC CAGTCCCTGT GCCTGGACTG 1440  
 ATTGGAACAC TTTGTTTTTC TCCCTGTGCC ATTTACCCCT CCACCTTTCC ATCCTGCCTT 1500  
 5 CTACCACCCT TGGATGAATG GATTTTGTA TTTCTAGCTGT TGTATTTTGT GAATTTGTTA 1560  
 ATTTTGTTGT TTTCTGTGA AACACATACA TTGGATATGG GAGGTAAAGG AGTGTCCCAG 1620  
 10 TTGCTCCTGG TCACTCCCTT TATAGCCATT ACTGTCTGT TCTTTGTAAC TCAGGTTAGG 1680  
 TTTTGGTCTC TCTTGCTCCA CTGCAAAAAA AAAAAAA 1717

15

(2) INFORMATION FOR SEQ ID NO: 125:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 804 base pairs  
 20 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 125:

CCACGGTCC GGTCACTATG TAGTGGAGGG GCAGACACCC TCCCGCAAAT TCTGGAAGGT 60  
 TCTTAGTCTC GACTAGGGCA GTAGCCCCAG GACTCCTAGT CGCGGGCTTC AGGTCACTGC 120  
 30 CGGCTGAACG GAGCTGCCGT CGCCATGTTT GGCTGCTTGG TGGCGGGGAG GCTGGTGCAA 180  
 ACAGCTGCAC AGCAAGTGGC AGAGGATAAA TTTGTTTTTG ACTTACCTGA TTATGAAAGT 240  
 ATCAACCATG TTGTGGTTTT TATGCTGGGA ACAATCCCAT TTCCTGAGGG AATGGGAGGA 300  
 35 TCTGTCTACT TTTCTTATCC TGATTCAAAT GGAATGCCAG TATGGCAACT CCTAGGATTT 360  
 GTACGAATG GGAAGCCAAG TGCCATCTTC AAAATTTCAG GTCTTAAATC TGGAGAAGGA 420  
 40 AGCCAACATC CTTTGGAGC CATGAATATT GTCCGAAGTC CATCTGTTC TCAGATTGGA 480  
 ATTTCACTGG AATTATTAGA CAGTATGGCT CAGCAGACTC CTGTAGGTAA TGCTGCTGTA 540  
 TCCTCAGTTG ACTCATTAC TCAGTTCACA CAAAAGATGT TGGACAATTT CTACAATTTT 600  
 45 GCTTCATCAT TTGCTGTCTC TCAGGCCAG ATGACACCAA GCCCATCTGA AATGTTTATT 660  
 CCGGCAAATG TGGTTCTGAA ATGGTATGAA AACTTTCAAA GACGACTAGC ACAGAACCCT 720  
 50 NTMTTITGGN AACATAATT TGAATAAAAT AATTTTAAAT GGATTMTGNA AAAAAAAAAA 780  
 AAAAAAAAAA AAAAAAAAAA AAAA 804

55

(2) INFORMATION FOR SEQ ID NO: 126:

(i) SEQUENCE CHARACTERISTICS:  
 60 (A) LENGTH: 431 base pairs

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 126:

	GGCACAGCCC AGGGCCTTGA AGCCAGCTGG CCCTGGAGAG GGGCTGCTGT GCCAGCTTGG	60
10	GGAGGGTCTG GGATGGGGCT GCCCCTGATG GCCCTGATGT GGAGTACCTT GCCAGCATCT	120
	GCTGGGGTGA ACTTTATTTT AGCCCTTCCC TTGTTGCTCT TATGGAAGAA CAGAGGAGGG	180
	GTGGGCAGGT CAGTGATGTC AGCAGTGGAG TGATTCCCAG CACAGCGGCT TCTGGGAAGA	240
15	GGGCATGGAG GCATTTCTTT CAGGGAAATG GTCCATNATT TCAGCCAGAA GGCATTGCAT	300
	TAAGTTAAGT CCNGGACTTT TGTGGCCAG CTCTGTGTTA TTAAGGGCCC TTGGCGAAGA	360
20	CTTCAAGGAG GGGGCAAAAN GACCTTTAAG TTTTtaggT TAACACAGGG AACCCNCAA	420
	GGGTTATTTT G	431

25

(2) INFORMATION FOR SEQ ID NO: 127:

(i) SEQUENCE CHARACTERISTICS:

30	(A) LENGTH: 3752 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
	(D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 127:

	NGGCACGAGG AGAGTCACCT GGA CTCAGAA CTAGAGATAT CCAATGACCC AGACAAAATT	60
	AAACTTCAGC TTTCTAAGCA TAAGGAGTTT CAGAAGACTC TTGGTGGCAA GCAGCCTGTG	120
40	TATGATACCA CAATTAGAAC TGGCAGAGCA CTGAAAGAAA AGACTTTGCT TCCCGAAGAT	180
	ASTCAGAAAC TTGACAATTT CCTAGGAGAA GTCAGAGACA AATGGGATAC TGTTTGTGGC	240
45	AAGTCGTGG AGCGGCAGCA CAAGTGGAG GAAGCCCTGC TCTTTTCGGG TCAGTTCATG	300
	GATGCTTTGC AGGCATTGGT TGA CTGGTTA TACAAGGTGG AGCCACAGCT GGCTGAGGAC	360
	CAGCCCGTGC ACGGGGGACC TTGACCTCGT CATGAACCTC ATGGATGCAC ACAAGGTTT	420
50	CCAGAAGGAA CTGNGAAAG CGAACAGGAA CCGTTCAGGT CCTGAAGCGG TCAGGCCGAG	480
	AGCTGATTGA GAATAGTCGA GATGACACCA CTTGGGTAAA AGGACAGCTC CAGGAAGTGA	540
55	GCACTCGCTG GGACACTGTC TGTAACCTCT CTGTTTCCAA ACAAAGCCGG CTTGAGCAGG	600
	CCTTAAACA AGCGGAAGTG TTTGAGACA CAGTCCACAT GCTGTTGGAG TGGCTTTCTG	660
	AAGCAGAGCA AACGCTTCGC TTTGGGGAG CACTTCCTGG ATGACACAGA GGCCCTGCAG	720
60	TCTCTCATTG ACACCCATAA GGAATTCATG AAGAAAGTAG AAGAAAAGCG AGTGGACGTT	780

	AACTCAGCAG TAGCCATGGG AGAAGTCATC CTGGCTGTCT GCCACCCCGA TTGCATCACA	840
5	ACCATCAAAC ACTGGATCAC CATCATCCGA GCTCGCTTCG AGGAGGTCCT GACATGGGCT	900
	AAGCAGCACC AGCAGCGTCT TGAAACGGCC TTGTGAGAAC TGGTGGCTAA TGCTGAGCTC	960
	CTGGAAGAAC TTCTGGCATG GATCCAGTGG GCTGAGACCA CCCTCATTC ACGGGATCAG	1020
10	GAGCCAATCC CGCAGAACAT TGACCGAGTT AAAGCCCTTA TCGCTGAGCA TCAGACATTT	1080
	ATGGAGGAGA TGAATCGCAA ACAGCCTGAC GTGGACCGGG TCACCAAGAC ATACAAAAGG	1140
15	AAAAACATAG AGCCTACTCA CGCGCCTTTC ATAGAGAAAT CCGCAGCGG AGGCAGGAAA	1200
	TCCCTAAGTC AGCCAACCCC TCCTCCCATG CCAATCCTTT CACAGTCTGA AGCAAAAAAC	1260
	CCACGGATCA ACCAGCTTTC TGCCCGCTGG CAGCAGGTGT GGCTGTTAGC ACTGGAGCGG	1320
20	CAAAGGAAAC TGAATGATGC CTTGGATCGG CTGGAGGAGT TGAAAGAATT TGCCAACTTT	1380
	GACTTTGATG TCTGGAGGAA AAAGTATATG CGTTGGATGA ATCACAAAAA GTCTCGAGTG	1440
25	ATGGATTTCT TCCGGCGCAT TGATAAGGAC CAGGATGGGA AGATAACACG TCAGGAGTTT	1500
	ATCGATGGCA TTTTAGCATC CAAGTTCCCC ACCACCAAGT TAGAGATGAC TGCTGTGGCT	1560
	GACATTTTCG ACCGAGATGG GGATGGTTAC ATTGATTATT ATGAATTGTG GGCTGCTCTT	1620
30	CATCCCAACA AGGATCGGTA TCGACCAACA ACCGATGCAG ATAAAATCGA AGATGAGGTT	1680
	ACAAGACAAG TGGCTCAGTG CAAATGTGCA AAAAGGTTTC AGGTGGAGCA GATCGGAGAG	1740
35	AATAAATACC GGTTCCTCCT CGGCAATCAG TTTGGGGATT CTCAGCAGTT GCGGCTGGTC	1800
	CGTATTCTGC GCAACCGTGA TGGTTCGGT TGGTGGAGGA TGGATGGCCT TGGATGAATT	1860
	TTTAGTGAAA AATGATCCCT GCCGAGCAG AGGTAGAAGT AACATTGAAC TTAGAGAGAA	1920
40	ATTCATCCTA CCAGAGGGAG CATCCAGGG AATGACCCCC TTCCGCTCAC GGGGTCGAAG	1980
	GTCCAAACCA TCTTCCGGG CAGCTTCCCC TACTCGTTCC AGCTCCAGTG CTAGTCAGAG	2040
45	TAACCACAGC TGTACATCCA TGCCATCTTC TCCAGCCACC CCAGCCAGTG GAACCAAGGT	2100
	TATCCCATCA TCAGGTAGCA AGTTGAAACG ACCAACACCA ACTTTTCATT CTAGTCGGAC	2160
	ATCCCTTGCT GGTGATACCA GCAATNAGTT CTTCCTCGGC CTCCACAGGT GCCAAAATA	2220
50	ATCGGGCAGA CCCTAAAAAG TCTGCCAGTC GCCCTGGGAG TCGGGCTGGG AGTCGAGCCG	2280
	GGAGTCGAGC CAGCAGCCGG CGAGGAAGTG ACGCTTCTGA CTTTGACCTC TTAGAGACGC	2340
55	ATTGCTTGTT CCGACACTTC AGAAAGCAGC GCTGCAGGGG GCCAAGGCAA CTCCAGGAGA	2400
	GGGCTAAACA AACCTTCCAA AATCCCAACC ATGTCTAAGA AGACCACCAC TGCCCTCCCC	2460
	AGGACTCCAG GTCCCAAGCG ATAACACTGT CTAAGCAGCC CCAAGCCACT ATCCACTTTG	2520
60	AATCCTGCTC CATACATTGG GTGTATATTT ATTCTGAACG GGAGAAGTTA TATTGTTAAA	2580

	AGTGTAAG AATAATTGTG TTATGAAGCT GCCTTATTTT TTTCTTTTT GTAAGTTACT	2640
5	ATTTTCATGT GAATATTTAT GTAGATAAAA TTGCCTCCT GGTAACCCCTG TAATGGATGG	2700
	GGCCAGAAA TGAAATATTT GAGAAAAACA AGTGAAAAGG TCAAGATACA AATGTGTATT	2760
	AAAAAAAAA AAGCCTATTA ATAGGGTTTC TCGCGGTGC AGGGTGTAA ACCTGCTTTA	2820
10	TCTTTTAGGA TTATTCCTAA ATGCATCTTC TTTATAAACT TGACTTGCTA TCTCAGCAAG	2880
	ATAAATTATA TTAATAAAT AAGAATCCTG CAGTGTATA GGAATCTTT TTTGTAAAT	2940
15	CACGACACC TCAATTAGCA AGAAGTGGG GGAGGGCTTT TTCCATTGTT TAATGTTTG	3000
	TGATTTTAG CTAAAGAGAG GGAACCTCAT CTAAGTAACA TTGCACATG ATACAGCAA	3060
	AGGAGTTCAT TGCAATACTG TCTTTGATA TTGTTTCAGT ACTGGGTGTT TAAAGGACAA	3120
20	ATAGCTGCTA GAATTCAGGG GTAAATGTAA GTGTCAGAA AACGTCAGAA CATTTGGGGT	3180
	TTTAACTGA TTTGTGCTC CCTATCCAGC CTAGACACCA GTAATCTTG TGTTCACCAG	3240
25	GACCCAGACC CTGGCAAGG GATAGGCTCG TTGGTGACAT TGTGAATTC AGATTGTTT	3300
	TATCCACTTT TTTGCTATT TATTTAAATG GTCGATCAAC TTCCACAAA CTGAGGAATG	3360
	AATTCACGA GCCTGTTCTG AAAATGTGGA CGTAAGACAA ACACGTGCTC GTCCTTTAAT	3420
30	GGAGTTCACC AGCAGACTTG TTAACAGTC CTGTTTGCTT TGTCTTTTT TTGTGCGTAA	3480
	TAAAGTCAAC TGACCAAGTG ACCATGAAAA GGGGCTGTCT GGGGCTCCTG TTTTITAGCT	3540
35	GCTGTCTTC AGCTCCGACC ATGTTGCTGT GTGATTATCT CAATTGGTTT TAATTGAGGC	3600
	AGAACTGAA GCTCTACCA TGAAGTGT TGAACAAGA CACTTTTG TATTAAATT	3660
	GCTGCAGTA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AACTCGAGG GGGGCCCGT	3720
40	ACCCAATTCG CCGTATATGA TCGTAAACAA TC	3752

45 (2) INFORMATION FOR SEQ ID NO: 128:

(i) SEQUENCE CHARACTERISTICS:

- 50 (A) LENGTH: 1144 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 128:

55	TGACCTCTG CCTGCCGGC TCAGTCTGG ACGCTTCTG TTTGTGCA GTCGGTCTC	60
	GGTAACACCA GCGCCTGTG GTCCACCACT CCATTCAGCA GCTCCATTG GTCCAGCAAC	120
60	CTAGCAGCG CCTTCCCTTC ACCACTCCAG CAAACAGCT GGCAAGCATC GGCCTCATGG	180

	GCACAGAAAA CTCCCCTGCT CCTCACGCTC CCTCCACCTC CAGTCCAGCT GACGACTTGG	240
	GACAGACCTA CAACCCGTGG CCGATATGGA GCCCCACGAT TGGAAGAAGA AGCTCGGACC	300
5	CTTGGTCTAA TTCGCACTTT CCTCACGAGA ATTAAATTAA GCAAAAAACA AACAAACATA	360
	GTGGGCCCTC GTCTAGATCA TGATGTGCCA GTTCTGAGA CATCTTTTTA AGGCTCTTAC	420
10	TGCAGCTCCC CTCCCCACCC TCCTCTTCTT TGCAAAACAG ACCCAAGCAG GGCAGGCTCA	480
	GACCACTCGC TTCTTTCAGA TCTTCTTTCG AATTATGATA ACATGAGATT TGCTGTGTG	540
	CTTTTAGAGA AAAGTCTGGA CTCAGCCACA AACTCTAATA AGACCTGTAC ATCTGAGAAC	600
15	CTTTCCCGTT ACTGCGTTTT CACCACCTGT CTTCCTCATG CTTTATTTAT CTGTATGAAC	660
	ACAGATTTGA CATTACAGCT AAGGAAATAA TTTGAGTTGA TTCAGAAATC CTGGCATGTG	720
20	ACAATTTTGT TAAATTACCA AGTTTGGTTT TTAATAATTT CTCAATATTA TGCGCCAAGA	780
	TCTAATTTTA AACTGTATG AGGACTTTGT GCTGAAAATA GAGTATTTTT TTAAAGTAAG	840
	GCTGTCTTGG TTTAAAGCA GATTACAGAA ATGTAAGTCA ACTTAAGAAC RGTGAATGAA	900
25	TGTAAAAACA TTCAGTYGAG ACCATATGCA TTTCTGTGC TGTTGTACT TGAGGTATGT	960
	AACATTTGTA TACCTGAAC TATTTTAAAG ATGAAGTAA ATGCACATAG CCAAGTCTTG	1020
30	AGATACAAGA TTGAATGTGT ATTTCTTAAA AATACAACTT TGTGTGTAC TTTGAAATAA	1080
	ATGATGCTTT TTTCAAAAAA AAAAAAAAAA AAAAAAAAC TCGAGGGGGG GCCCGGTACC	1140
	CAAT	1144

35

## (2) INFORMATION FOR SEQ ID NO: 129:

40

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1830 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

45

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 129:

	GCATGCAGAG GAGCACCTG AGCGTGTGCC TGGAGCAGGC GGCCATSTTG GCACGGAGCC	60
50	ACGGGTGCT GCCCAAGTGC ATCATGCAGG CCACGGACAT CATGCGGAAC AGGGCCCAAG	120
	GGTGGAGATT CTGGCCAAAA ACCTGCGAGT CAAGGACCAG ATGCCCCAGG GTGCTCCGCG	180
55	CCTCTACCGC CTCTGCCAGC CGCCGGTGGG TGGGACCTC TGAACACCCA AATGCCCCAC	240
	GCTGGGCCGC GGCTCTGGA GCTGGGATT GGGAGGACAC AGCAGGCAGC GCTGGCCTTC	300
	TCCAGGGATG GCCCAANGCT TCCGCARCCG CCCGTTCCGG GACCTGCCCA GCGTCTCCC	360
60	TGCTCTCTC CGGGACAAGC CTGGCCACCC TCGCTGTGAT GACGAGCTGG CTGATGGCC	420

	CTGGGCCGGC CCATTCTTCA CACGCCCTGCC AGAAGCTGGA GGGGTGCTGG AGACCCATAG	480
5	AGCTGATGGG AGCAGCTGGT GCCTGGCCTT OGGCTCCTGC GTCCCCAGAA CCCAAGGGAA	540
	CGTCATGGAG GCCACATGGG GCCACCCGGC TCCCTCGGGA TGGCTCCGCT GCACTTTTGA	600
	AACCCCGGTT TCCTTCAACG TCCACATTCC AGGTGACCAC ACGTGTCTCC TCCTCCTCAT	660
10	CTTAGCTTCC AGGTTACACC TAACCCCTGTA CTAACCTGCT TGGTGGACTT GGAAAAGACT	720
	TGGCTCTGTC GGGAAAGGAG AGACGGGGCC TCCATCACGC CTGTTACCAG AGGATCCCCG	780
15	AGAGCCACAC CAGCTCTGGA CATCACCGCC CCTGGAACTG GGGCCACCAG CCCTGGGCAC	840
	GAGATTTGCT CTGACTTTAT TTATATGGCA TGAAATCTCT GGTTTATTTT GGGATTTTTT	900
	GTTGTTGGTG TTGTCAAAGT TTGTTTMTTC TAAAGTTGTG TGATTATATA TTTGACATTT	960
20	TACATTTCOA AGAAAGGTAT GTGTCTAAC AGGGGACCAA CAGAAGGTAG TATTGACAAC	1020
	TGTTCTCTCT TCTACTAAAA AAAAAAGAGC ACAAAGAAA AACTAAATTA TTGAAAAATT	1080
25	AAAAAATGTC ATTGTTTCCT GTTGTGTAAT ATTAGGGTGT TAAGGTGTCT TTTTGAGGTA	1140
	TCGACTGTGA TTCCTTCCCC CACCCTCCAT TCTCCAGCGG TTGGCCGGTG TTAGAACTCG	1200
	CTCTCTTTGA GTGACTGGCT ACAAGGGCCT GAGAGGTGGC CAGCCAGGCT TGGAGCTGGA	1260
30	GGGGATGGAG CCCACCTGA GGTGCCGTGT CACACGGGTT AGAGGGTCAC TGGGAAACAC	1320
	CGGGCGGTGG CTTCTGTGAT TTATTTTCTT GATGGTAACT TCTCAGAGCA GGGCRATTGG	1380
35	GACATCACCA GCCAGAGCAC AGGAAGCCAC CCTGCCTGCT GGGGAGGAGG GACCCACACA	1440
	AGCCCCCTCG GCAGTTTGTC CCCCAGCTT CGGTATGCCT TCAGGGAAAG GTCACAGCTG	1500
	GGGAGGAAGC GGGGGGACGC CTGTCACCCC TGGCAGGTGG TGAGTTCAGG TGGGGGCTCC	1560
40	CTGCTKCCCC CAGGCCTGGG AGCTTGAAGC CCTCCCGGCA TCTGGCATCC GAGCCTCCCG	1620
	CCCTCCAGGG TGGGCTTCCC TCTCTTGCCG CAGCATACAC GAGGGCAGGC AGTGGCCTTG	1680
45	TCACTGTATC TTGCATCAGA GACAAAGGAG GACCCGCTTT AGCCCTGCTG CGGGAAATGG	1740
	GGGATGGCCC AGGGCCAGCG CATTGTGCAC TGGTTTACTT TAAAATGTAC AGATTCTTCT	1800
50	CGTTAAATTC TTGATAGATT TTTTATTATT	1830

(2) INFORMATION FOR SEQ ID NO: 130:

- 55 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1864 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- 60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 130:

	GGCCGCCCGG ATGGCGACCC CAGCCTCGGC CCCAGACACA CGGGCTCTGG TGGCAGACTT	60
5	TGTAGGTTAT AAGCTGAGGC AGAAGGGTTA TGTCTGTGGA GCTGGCCCCG GGGAGGGCCC	120
	AGCAGCTGAC CCGCTGCACC AAGCCATGCG GGCAGCKGGA GATGAGTTGG AGACCCGCTT	180
10	CCGGCGCACC TTCTCTGATC TGGCGGCTCA GCTGCATGTG ACCCCAGGCT CAGCCCAACA	240
	ACGCTTCACC CAGGTCTCCG ATGAACCTTT TCAAGGGGGC CCCAACTGGG GCCGCTTGT	300
	AGCCTTCTTT GTCTTTGGGG CTGCACTGTG TGCTGAGAGT GTCAACAAGG AGATGGAACC	360
15	ACTGGTGGGA CAAGTGCAGG AGTGGATGGT GGCCTACCTG GAGACGCGGC TGGCTGACTG	420
	GATCCACAGC AGTGGGGGCT GGTATCCCA GATCACTGAA GCTGAGATGG CTGATGAAGT	480
20	AATTTGCAGT GAAATTTTAA GCGACTGTGA CTCTGCTGCA AGTTCCTCCAG ATCTTGAGGA	540
	GCTGGAAGCT ATCAAAGCTC GAGTCAGGGA GATGGAGGAA GAAGCTGAGA AGCTAAAGGA	600
	GCTACAGAAC GAGGTAGAGA AGCAGATGAA TATGAGTCCA CCTCCAGGCA ATGCTGGCCC	660
25	GGTGATCATG TCCATTGAGG AGAAGATGGA GGCTGATGCC CGTTCATCT ATGTTGGCAA	720
	TGTGGACTAT GGTGCAACAG CAGAAGAGCT GGAAGCTCAC TTTCATGGCT GTGGTTCAGT	780
30	CAACCGTGTT ACCATACTGT GTGACAAATT TAGTGGCCAT CCCAAAGGGT TTGCGTATAT	840
	AGAGTTCTCA GACAAAGAGT CAGTGAGGAC TTCCTTGGCC TTAGATGAGT CCCTATTTAG	900
	AGGAAGGCAA ATCAAGGTGA TCCCAAAACG AACCAACAGA CCAGGCATCA GCACAACAGA	960
35	CCGGGGTTTT CCACGAGCCC GCTACCGCGC CCGGACCACC AACTACAACA GCTCCCCTC	1020
	TCGATTCTAC AGTGGTTTTA ACAGCAGGCC CCGGGGTCGC GTCTACAGGG GCCGGGCTAG	1080
40	AGCGACATCA TGGTATTCCT CTTACTAAAA AAAGTGTGTA TTAGGAGGAG AGAGAGGAAA	1140
	AAAAGAGGAA AGAAGGAAAA AAAAAAGAA TAAAAAATA AAAAAAATA ACAGAAGWTG	1200
	MCCTTGATGG AAAAAAATA TTTTPTAAAA AAAAGATATA CTGTGGAAGG GGGGAGAATC	1260
45	CCATAACTAA CTGCTGAGGA GGGACCTGCT TTGGGGAGTA GGGGAAGGCC CAGGGARTGG	1320
	GGCAGGGGGC TGCTTATTC ACTTGGGGAT TCGCCATGGA CACGTCTCAA CTGCGCAACT	1380
50	GCTTGCCCAT GTTCCCTGC CCCACCCAC CCTCTTCTC CGGCTCCCTG CCCCTCCAGA	1440
	TTGCCCTGGT ATCTATTTTG TTTCCTTTTG TGTTCTTTT TCTGTTTGA GTGCTTTCT	1500
	TTGCAGGTTT CTGTAGCCGG AAGATCTCCG TTCCGCTCCC AGCGGCTCCA GTGTAAATTC	1560
55	CCCTTCCCCC TGGGGAAATG CACTACCTTG TTTTGGGGGG TTTAGGGGTG TTTTGTTTT	1620
	TCAGTTGTTT TGTPTTTTGG TTTTPTTNT TTTCCTTTGC CTTTPTTCCC TTTTATTTGG	1680
60	AGGGAATGGG AGGAAGTGGG AACAGGGAGG TGGGAGGTGG ATTTGTGTTA TTTTPTTAGC	1740

TCATTTCCAG GGGTGGGAAT TTTTITTTTAA TATGTGTCAT GAATAAAGTT GTTTTGTAAA 1800  
 AKAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1860  
 5 AAAA 1864

10 (2) INFORMATION FOR SEQ ID NO: 131:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2041 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 131:

20 GGCACGAGCG CGCGGCAGGG CCTTGGACCC GCGCGGCTCC CGGGGATGGT GAGCAAGGCG 60  
 CTGCTGCGCC TCGTGTCTGC CGTCAACCGC AGGAGGATGA AGCTGCTGCT GGGCATCGCC 120  
 25 TTGCTGGCCT ACGTCGCCCTC TGTITGGGGC AACTTCGTTA ATATGAGGTC TATCCAGGAA 180  
 AATGGTGAAC TAAAAATTGA AAGCAAGATT GAAGAGATGG TTGAACCACT AAGAGAGAAA 240  
 ATCAGAGATT TAGAAAAAAG CTTTACCCAG AAATACCCAC CAGTAAAGTT TTTATCAGAA 300  
 30 AAGGATCGGA AAAGAATTTT GATAACAGGA GCGCGAGGGT TCGTGGGCTC CCATCTAACT 360  
 GACAACTCA TGATGGACGG CCACGAGGTG ACCGTGGTGG ACAATTCTCT CACGGGCAGG 420  
 AAGAGAAACG TGGAGCACTG GATCGGACAT GAGAACTTCG AGTTGATTAA CCACGACGTG 480  
 35 TGGAGCCCTC CTACATCGAG GTTGACCAGA TATACCATCT GGCATCTCCA GCCTCCCTC 540  
 CAACTACAT GTATAATCCT ATCAAGACAT TAAAGACCAA TACGATTGGG ACATTAAACA 600  
 40 TGTITGGGCT GGCAAAACGA GTCGGTCCCC GTCTGCTCCT GGCTCCACA TCGGAGGTGT 660  
 ATGGAGATCC TGAAGTCCAC CCTCAAAGTG AGGATTACTG GGGCCACGTG AATCCAATAG 720  
 GACCTCGGC CTGCTACGAT GAAGGCAAAC GTGTTCGAGA GACCATGTGC TATGCCTACA 780  
 45 TGAAGCAGGA AGGCGTGGAA GTGCGAGTGG CCAGAATCTT CAACACCTTT GGGCCACGCA 840  
 TGCACATGAA CGATGGGCGA GTAGTCAGCA ACTTCATCCT GCAGGCGCTC CAGGGGAGC 900  
 50 CACTCACGGT ATACGGATCC GGGTCTCAGA CAAGGGCGTT CCAGTACGTC AGCGATCTAG 960  
 TGAATGGCCT CGTGGCTCTC ATGAACAGCA ACGTCAGCAG CCCGGTCAAC CTGGGAACC 1020  
 CAGAAGAACA CACAATCTTA GAATTTGCTC AGTTAATTAA AAACCTTGTT GGTAGCGGAA 1080  
 55 GTGAAATCA GTTCTCTCC GAAGCCCAGG ATGACCCACA GAAAAGAAAA CCAGACATCA 1140  
 AAAAAGCAA GCTGATGCTG GGGTGGGAGC CCGTGGTCCC GCTGGAGGAA GGTITAAACA 1200  
 60 AAGCAATCA CTACTCCGT AAAGAACTCG AGTACCAGGC AAATAATCAG TACATCCCCA 1260



	AACCAAAGCC TGCCAGAATA AAGAAAGGAC GGA CTGCGCA CAGCTGAACT CCTCACTTTT	1320
5	AGGACACAAG ACTACCATTG TACACTTGAT GGGATGTATT TTTGGCTTTT TTTTGTGTGTC	1380
	GTTTAAAGAA AGACTTTAAC AGGTGTCATG AAGAACAAAC TGAATTTCA TTCTGAAGCT	1440
	TGCTTTAATG AAATGGATGT GCCTAAAAGC TCCCCTCAA AACTGCAGA TTTTGCCTTG	1500
10	CACTTTTGA ATCTCTCTT TTATGTAAAA TAGCGTAGAT GCATCTCTGC GTATTTTCAA	1560
	GTMTTTTAT CTGCTGTGA GAGCATATGT TGTGACTGTC GTTGACAGTT TTATTTACTG	1620
15	GTTCCTTGT GAAGCTGAAA AGGAACATTA AGCGGGACAA AAAATGCCGA TTTTATTAT	1680
	AAAAGTGGT ACTTAATAAA TGAGTCGTTA TACTATGCAT AAAGAAAAAT CCTAGCAGTA	1740
	TTGTCAGGTG GTGGTGGGCC GGCATTGATT TTAGGCAGA TAAAGAATT CTGTGTGAGA	1800
20	GCTTTATGTT TCTCTTTTAA TTCAGAGTTT TTCCAAGGTC TACTTTTGAG TTGCAAACTT	1860
	GACTTTGAAA TATTCTGTG GGTGATGATC AAGGATATTT GAAATCACTA CTGTGTTTGT	1920
25	CTGCGTATCT GGGCGGGGG CAGGTGGGG GGCACAAAGT TAACATATTC TTGGTTAACC	1980
	ATGGTTAAAT ATGCTATTTT AATAAAATAT TGAAACTCAC CAAAAAAAAA AAAAAAAAAA	2040
	A	2041

30

(2) INFORMATION FOR SEQ ID NO: 132:

35

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2012 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 132:

	TACCAAGCTG CAAGAATCTA CTATATCATG GCAGAAGAAG TAGAGTGGGA CTATTGCCCT	60
45	GACCGGAGCT GGAACGGGA ATGGCACAAC CAGTCTGAGA AGGACAGTTA TGGTTACATT	120
	TTCTGAGCA ACAAGGATGG GCTCCTGGT TCCAGATACA AGAAAGCTGT ATTCAAGGAA	180
50	TACACTGATG GTACATTGAG GATCCCTCGG CCAAGGACTG GACCAGAAGA ACACTTGGGA	240
	ATCTTGGGTC CACTTATCAA AGGTGAAGTT GGTGATATCC TGACTGTGGT ATTCAAGAAT	300
	AATGCCAGCC GCCCTACTC TGTGCATGCT CATGGAGTGC TAGAATCTAC TACTGTCTGG	360
55	CCACTGGCTG CTGAGCCTGG TGAGGTGGTC ACTTATCACT GGAACATCCC AGAGAGGTCT	420
	GGCCCTGGC CAATGACTCT GCTTGTGTTT CCTGATCTA TTATTCTGCA GTGGATCCCA	480
60	TCAAGGACAT GTATAGTGGC CTGGTGGGC CCTTGGCTAT CTGCCAAAG GGCATCCTGG	540

	NAGCCCCATG GAGGACGGAN TGACATGGAT CGGGAATTG CATTGTTGTT CTGATTTTT	600
	GATGAAAATA AGTCTTGGTA TTTGGAGGAA AATGTGGCAA CCCATGGGTC CCAGGATCCA	660
5	GGCAGTATTA ACCTACAGGA TGAAACTTTC TTGGAGAGCA ATAAAATGCA TGCAATCAAT	720
	GGGAAACTCT ATGCCAACCT TAGGGGTCTT ACCATGTACC AAGGAGAAG AGTGGCCTGG	780
10	TACATGCTGG CCATGGGCCA AGATGTGGAT CTACACACCA TCCACTTTCA TGCAGAGAGC	840
	TTCTCTATC GGAATGGCGA GAACTACCGG GCAGATGTGG TGGATCTGTT CCCAGGGACT	900
	TTTGAGGTTG TGGAGATGGT GGCCAGCAAC CCTGGGACAT GGCTGATGCA CTGCCATGTG	960
15	ACTGACCATG TCCATGCTGG CATGGAGACC CTCTTCACTG TTTTCTCTCG AACAGAACAC	1020
	TTAAGCCCTC TCACCGTCAT CACCAAAGAG ACTGAAAAAG CAGTGCCCCC CAGAGACATT	1080
	GAAGAAGGCA ATGTGAAGAT GCTGGGCATG CAGATCCCCA TAAAGAATGT TGAGATGCTG	1140
20	GCCTCTGTTT TGGTGGCCAT TAGTGTCAAC CTCTGCTCG TTGTTCTGGC TCTTGGTGGA	1200
	TGCGTTTGGT ACCAACATCG ACAGAGAAAG CTACGACGCA ATAGGAGGTC CATCCTGGAT	1260
25	GACAGCTTCA AGCTTCTGTC TTTCAAACAG TAACATCTGG AGCCTGGAGA TATCCTCAGG	1320
	AAGCACATCT GTAGTGCAT CCCAGCAGGC CATGGACTAG TCACTAACCC CACACTCAAA	1380
30	GGGGCATGGG TGGTGGAGAA GCAGAAGGAG CAATCAAGCT TATCTGGATA TTTCTTTCTT	1440
	TATTTATTTT ACATGGAAAT AATATGATTT CACTTTTCTT TTAGTTTCTT TGCTCTACGT	1500
	GGGCACCTGG CACTAAGGGA GTACCTTATT ATCCTACATC GCAAATTTCA ACAGCTACAT	1560
35	TATATTTCCT TCTGACACTT GGAAGGTATT GAAATTTCTA GAAATGTATC CTCTCACA	1620
	AGTAGAGACC AAGAGAAAAA CTCATTGATT GGGTTTCTAC TTCTTTCAAG GACTCAGGAA	1680
40	ATTTCACTTT GAACTGAGGC CAAGTGAGCT GTTAAGATAA CCCACACTTA AACTAAAGGC	1740
	TAAGAATATA GGCTTGATGG GAAATGAAG GTAGGCTGAG TATTGGGAAT CCAAATTGAA	1800
	TTTGTATTCT CCTGGCAGT GAACTACTTT GAAGAAGTGG TCAATGGGTT GTGCTGCCA	1860
45	TGAGCATGTA CAACCTCTGG AGCTAGAAGC TCCTCAGGAA AGCCAGTTCT CCAAGTTCTT	1920
	AACCTGTGGC ACTGAAAGGA ATGTTGAGTT ACCTCTTCAT GTTTTAGACA GCAAACCCTA	1980
50	TCCATTAAAG TACTTGTTAG AACACTGAAA AA	2012

(2) INFORMATION FOR SEQ ID NO: 133:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1669 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 133:

5	GAGCAGTATT TTAACCAACT TGTATTACAG ATGTTACAGT TCATGTTAGG AAGTCAGAAA	60
	AGACTTTGTT TGTCTTTGTT CTGCTGATGT GAGTCATGTT TTGTGGGGTC TTCCATGGCA	120
	CATTTACCTG TTGCTCCGTC CAGATGTTGA GGGCCAGTCT AGGCTGACAC ATCCTACCCG	180
10	AGGACAAGCC TGTCTCAT TCTTCACTC TCCCCTCCCC ATATAGCAAC TCTCCCAGGT	240
	TTAGATTACC GTTTTCGACG ACAGATTAAC CAAAAATGCC CCACACAGGT TTTATTACTG	300
15	TTATATACTA TACTTTTAAC AGTACAGACC CTAAATTTTA TTATTTGTG CTCCCCCAAT	360
	CTGATACCAA ATGTTTAAAG TTGTTTGAAA TCCAAACATG GTAGTGTCA TGGGTAAATA	420
	TTTTCTAGGC TATGTAAGAG TTAGCAGCCC ATAGCATAGA AGTAATCAAG TAGCATCTGA	480
20	GACTGTTGGA GGCAC TAGGG CCTCTCTGGG CCTAACAGCC TCACTTCCCC AGCCTCACCT	540
	TGCTGTCCCTC TGACACTGCC ATCAGGGCTG TTAGTGGCAC CTGTATGAGG CCAAGTGTGC	600
25	GTCCAGGGGA ACAGCACAGG TTAATGCGTC TCCCTAGAAC TCATGAAGTC AGTTTAATTC	660
	ATGCATGAAC ATGAGTTCAT TTTATGTTTT ATATAGCTTT CTAGACATA CCAAACCATC	720
	ATTCATAAAT CAGATAAATT ATTCAGTTTT TGTGTTTGA AAGCTAAGTA TGTGTAGCTG	780
30	GAAACAAAAA TGAGCGTGT TCTCTCCTG TTAATCTAGA GTGTGCAGTT ACACATGTGT	840
	GGATAATTTT ATGTTCCAGG GCGCTTGGC ATCTCCCATG GACTGATTCC CAGGAAGAAA	900
35	AGCCCAAAGG GAAACCCACG ATTCTTTTCG AGTAGATGTG GGAAAGAGCC CATTGGAGGA	960
	TATGAGGTCC TGTGAAATTC AGTGTGTGT GTGGCTCCTT GTTAGCAGTC ATGTTGACAT	1020
	GGTGTTAGGA GGCTCCCCAT CCACCTTTTA CATGATGTAG GGACCAGTGT CTTGTGAGAT	1080
40	TAACCTTGGG ACACAGTGGG TTAGCCTGGA GAAAATGAGA GGCCCTGCCT GGACCCAGGG	1140
	AGAGGAGCCA GTGACACAGG CAGAGCGGTG CAGCCCTCCT TCCCTTCCAT TTGGAGGAGG	1200
45	TGGTGCCAGG AGCCTGCCCG CTTACCTCTG CTGAAGCATA AGTGGACTTT GCTTTTGGGG	1260
	CTTATCTCTG ATACATGCTG GAGCCCTGCC TCTCCACTGC TAGATGGAAC CTGGAATCTC	1320
	TCATCTACCT CTTAGTCTGT CAGTTTCTAC GTGTGAGAAG CAAGCTGTG GGCCAGTGTG	1380
50	CTTGATACATG CTGTAGCACT TAAAAATAA TTCCAGGGTT CCTGGAAAA CCAGTCCCAG	1440
	GGTTCTATG ATCTGTAGTT TCTACCTGGA TTATAACTGG TTTTGGGTAC CTGAATTTTG	1500
55	ATTGTTAGC CTTAATATA GTCTGGCGTG ATCATGTAGA ATCTTTCTG GTGAACAGAT	1560
	CATAAAGTTC TATCAAGGAG TTCTATCAAG GCATCCATGT CAGTGGTGCT ATGCTGGTTA	1620
	CAACTTGAGA TTTTGTAAAT AAAAAATTG TCATAAAAAA AAAAAAAAAA	1669

## (2) INFORMATION FOR SEQ ID NO: 134:

5

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1565 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 134:

	CACTTTGTCT ATATAACCTA AGTGATAACC CTCTTTTAGT TACCTGCCAA ACTCTGGNCT	60
15	TGGTTTATAT TGCAGTTAAC ACAGTTACAA AGCTGTAATG GTGTCTTTTT TTCCTTTGTA	120
	ACGGAATGTG TAAATCAAAG TATATACATT GTGTGGTGT CCGTTTCTG GAGTTTCATG	180
20	AGGATTTACA CATGGCATT CAGTGTCTGT ATAGATCTGC CTACCTTTGT GAATTCATCT	240
	GTTAACCCCT CTTCCTTTGA GAGAGCACCG GCGATGGTGG TTAACCTCTT GTGTTTCTC	300
	TCTCTCTTAC TGGTTATTCT TGAATTAAGC ACAGACTCGT CAGCTCGGTT GCTTTATCAT	360
25	GAATAATGTG TGTGACCTTG CAGTCTCTCC ACAGTTCAGC AAACAAGTGC TAGCTTCACT	420
	GACCAAAAT TAAGGAAGGA AAACACAGTT TTTAAAACGA TCCATCTTTT AACAGCCGAA	480
30	ACCGATGTGT CTATGGTGCT GCACCTTGCT GTTGTAATTC TGAAATCAGA CGTGTGTGAA	540
	CGATCATTTT TGACTTAACC GTGAGATGCT CACGAGTACC CTTCCTGTTG TTTTGTTAGC	600
	ATTGAAATCG AGACTATTTA TTTGGAATAT ATACAACAGT GTTTTCCAC TGTATTTTAT	660
35	TTGCAAAAGT TGAGAACTGC TTTCTCTACC TTTTGCAAAA TAATTGATAT TCCATATTGG	720
	ATTCTCAAAG ACTTCGATAT GGTGAACCTA TTAACCTAG AAATTGTATT CATCCTTTCA	780
40	TGACTGTGGC CTGAGTTCCC CAGCCCCCTCT CCTCCTTTT TTTAGATGAG ATTTAGCACA	840
	CTCTCAGTTA TTTAAACATG CAACATTTCT TGAGTATGTA TGTGAGGCC ATCTGAGCTC	900
	ATAGCTGATT CAGTAACCAG TTTTCATGCTG TGTCAATCAC ACTCACTACT TAATACTGCC	960
45	ATGGTGAAAA TGTGGAGGAA AAATGTATCC ATGTGTGTCT GGAAGCATA TACACTTGTA	1020
	CATTTTTTAA TACTCTGATT CTGTAACATT TCTGAGTTT GTTTTGT TTTT ACAGNAAAAA	1080
50	AAAAAAAAGT GATAAAGCAA TCAGAAGACC AAGAGGTTTA CTATTGATGC TTAGGGTCGT	1140
	CTGACCTTGG CTGGCCAATA GACCTACACG GCCAAATTAA TTTACGAGAG TAATAATTTT	1200
	TCAAAGCCA ATTTTTTTT TGTATTTTCT GTATGAACT GCCAATATCA TGAATAGAAA	1260
55	GGGAGAACCA TAAAGGAGAA AGAACGTGAT GTTCTGTTAT GTTCATGTAA ACCTAAAGAA	1320
	ACAGTGTGGA GGCAGGCGCG ATCAGCCGAA CTCTAGGGAC TTGGTGTGTC TTGGAAGGCA	1380
60	TCCATACCTG CATTTTGCAT TCTTCGTATG TAATCATATT GCCAAAGACA AACTATTTCA	1440

	TCATTTATTG TAAATAACAC TTTTCCCCAG ACCTACCATA AAGTTTCTGT GATGTATTGT	1500
	CTTCCAGTTG CAATAAAAT TACTGAGTTG CATCAATTGA AGAAAAAAA AAAAAAAA	1560
5	CTCGA	1565
10	(2) INFORMATION FOR SEQ ID NO: 135:	
	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 2007 base pairs	
	(B) TYPE: nucleic acid	
15	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 135:	
20	TCTAAAAGCC CCTTATACC CCACTTTGTG CAGCAAAGAT CCCCCTGCAG GTCACAGCCT	60
	GATTGTGGC CAGGCTGGAC AAATTCCTGA GGCACAACTT GGCTTCAGTT CAGATTTCAA	120
	GCTGTGTTGG TGTGGGACC AGCAGAAGGC AAACGTCCAG CCAACACACA GGACTGTAAG	180
25	AGGACTCTGA GCTACGTGCC CTGTGAAGAC CCCCAGGCTT TGTATAGGA GGTCGTTTCA	240
	CTTCCCCAAA GTCAGAGGTG ATTTGATTG GGAAGACTG AATATTACA CCTAAGTCGT	300
30	GAGCATATCC TGAGTTTAC TTCCTTATGG CTGCCCCTCC AAGTTCTCTC TCTCATACAC	360
	ACACACACCC TTGCTCCAGA ATCACCAGAC ACCTCCATGG CTCCAGCTAT GGAACAGCT	420
	GCATGGGGC TGCCCTTCTG TTTGGCTTAG GAACTTCTGT GCTTCTTGTG GCTCCACTCG	480
35	CGAGGCAGCT CGGAGGTGTG GACTCCGATT GGGCTGCAGG CAGCTCTGGG ACGGCACAGG	540
	GCGGCGCTC TGATCAGCTC GTGTAAACA CACCGTCTC TTGGCCTCCT GGCAGTTCTT	600
40	TCTGCGAATA GTCCTCTCCC TGGCCAGTTG AATGGGGGAA GCTGCTGGCA CAGGAAGGAG	660
	AGGCGATCCC GGCTGAGGCT TAGGAAATG CTGGAGCCGG CTCCAAGCAG ATAATTCACT	720
	GGGAGGPTT TCAGAGTCAA ACATCATTTCT GCCTGTCTTG GGGCCAGGT GTGTCACACA	780
45	AGCATCTCAA AGTCAAAAGC CATCTGGGGC TGCTGCTTCT CTTTCTCAGG CTCTGGGGAA	840
	AGGAATCTCC CTCTCCTCTC ACTTGATTCC AAGTGTGGTT GAATTGTCTG GAGCACTGGG	900
50	ACTTTTTTTC TCTTTTCTT GATGGACCAA CAGTGCAAAT GCAATCTCGC CATTTAACTT	960
	TCAGGTCGAT TTCCTTTCCT GATCAGACAT CTTTGTGCCC CTTTTAGGAA GGAAAAGAAT	1020
	ACACCTACGA TGTGCCAGGC ACTGTGTTAG GCGCTTTTAT ATAGATCCTC GTTAGGATGA	1080
55	GAATAAGGGA TGAGGACATC TCTTTATAAA AGGCCCTAA GTAATGGATA AACAGAAACA	1140
	CTTAGAGGTG AGAAGGTCTG TCTTCAAGAT CCAAGGTAAG ATTGCCTTCA GTCTGATGTT	1200
60	TGTTCTCAAG GACTTATCCC CTACAATATT CTCCCACTCC ATACTTCTCC TTCTACCCCA	1260

5 CCATGTGCTC CCGTGCACTC CTCAGATGGT CAGAGGGGTA ACCCAAGTCC TTAGAGAATT 1320  
 TGGGGACCAA TAGAATATGT GATGTGTGAA TTTTCTTTAA AAAACTTAAG GAGTCTTTGC 1380  
 TACCTTCTGC TTGTTGAGTT GTTTTGGCAT TCATATTAAA AGCCAGCATC TCACTATTTA 1440  
 TTGACAGGTT GGGCTGTGTG TGTGCGCATG TGTGTATACA TTTCCAGGCG TGCCTGTGTC 1500  
 10 CTGTAGCTTT TTAAGAGGAA ACCCAGTCAT CCCACTATGA ATCTGGCATC TTCTTATGCT 1560  
 TCTAGTGTTC TGGCCATACA TCAACCAAGG GGTTTAATTT ATCCAATGCT TGACGACATG 1620  
 TTCAGGAGGG GCTGGATCAA ATTTTGAGAG GGTTATGGGA AAGGGAGGGG GAGAAGAAAT 1680  
 15 TGACATTTAT TTTATTATTT ATTTTAAATG TTTACATCTT CTTTATGTTG TATCAAGCCT 1740  
 GAATAGAAAC TGATAGCATT AAAATACTCC GTTCCTCTCT CTCTTCTCGC TTCCTTTTTT 1800  
 20 TTTTTTTTTA AATTTAGGAT AACACATTTT TGTTTCTAAA GTGATTTGTG ATTTGTGCTG 1860  
 TATAAACTGT ATAAAAGGTT CTGTTTTTAA AGGTGGATTT TCATTCTCTT GGGGACAGTG 1920  
 25 GTCGCCAAGA CATCTACATT GTAAGAGAAC ACAGTGAAG ATCCTGTCCT GATTCTCAAA 1980  
 AATTATTTTC TCTGTATGAT TAAAAGT 2007

30

(2) INFORMATION FOR SEQ ID NO: 136:

35 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1291 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 136:

CTTTTAACCC TCCCCCTTCA CACACATACA TATCAGGTTG TTTTCTAGTT AAAAACCCTA 60  
 GTAGCTCAGA TTCTACTTTA ATGTCAGTGC AGATTTCAT TGAATCATGC CATTATGTTT 120  
 45 TTTCTCATTT TTATGCTGTT GGGTCTTAGT TTTTAAATTG ATATAAGAA CTCAGCAATG 180  
 GTTTTATTTT CTACTCATAC TTAGGGTTTA GGAAACACTA CCACTAGTTA TCATTTAATC 240  
 AACTTCAATG GTCTACTGAA ACAAAAATGG TAACTTTTCA TTAGTGGATT ATTTAGAGTT 300  
 50 ATAGTAGTTG TTTCCAGAAA AACTTCTCTC ACAATGTGAC TTCCAATCA AATCATGTGA 360  
 TCATACAGTT ATTCCCATGA AAGGCAGAAAT GTTTGTTTCA AAATTAATCT AGTTTTCTGT 420  
 55 ACATTTAAAT TTGAGAAGGT GACAACGGC TCTTTTCCAG TCTTCTTCA GTTCAGTTTT 480  
 CTGATAGACC ACTATTGGCA AACAGTATCT GTCAACTACC AAATGTGTAA AATTTTCTGT 540  
 ATTTCACTTT GTCTTATTTG TAAATAGTGA ACTAAACTT TTGGCAGATC AGCAACATTT 600

60

	GCTGAGCCTG TTTTAAAGC TAATGTGTAT TCTTACTAAT GTTCCTATCA AGAATGGATT	660
	TGTAATATAT GCTGTCTATT TCTAATGTTT ACATTTCATAT TTGAGGTTT TATCTTATTT	720
5	TAATAGAGAA CAGACTTCTC AAAAAATCTT CAGAAGCAGC TTATTATTGA AATATCGAAA	780
	TATTGAAATA AACCCGGTGG GTTAGATTAC TCATCTGTCC ACCAAGTGGG ACATTTGCAT	840
10	GGACTGGGGG CTTAAAGGAC TTAGAAGAGA CCTGTAAGTA AATCCTGAAA ATGAGCCAAT	900
	CCCCACTTGA ATGTTTACTG GAGTAAACCC ACCTTTACCA CCCCAATTAC AGCACCCGAG	960
	GCCGATAAAC CAACTTGGCT CTGGTTCATT TTTCTTTTCT TCATTTGTGA TGCTCAGATT	1020
15	CAAAATGTGT GTTCTACACT GTTACAGGCT TCTCTTTTGT TTGATTAAAG ATTTTAGTCC	1080
	TACTTTTGTA TGGACACATT AGAATATTCA GAGACCAAAA TAGAAGAATT TGCTGTTAGA	1140
20	TATTTTTCAG AAGTCAGCAG ATTTGTGGCA AATCATTTAT TTGCCTTTT AAAAAATCAT	1200
	TTAAGCAGTT CAGAGAGTAG ACTACTCAGA AAATTATTTT ACGTAATTGT CTAAGAGGTC	1260
	AATATTTTTT AATGCATATT GAATCAAATA A	1291
25		

## (2) INFORMATION FOR SEQ ID NO: 137:

30

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1906 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

35

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 137:

	GGCAGGAGGA CCTACTTTTG TAACAGACCA TGGTTGTGTC CAAGGTAAAA CCACAGTGAT	60
40	ATTTTGGGAT GCTTTGTCTG CAATCTTGAC TTGTTTGTGC AGTATCATTA TTCAGACTTC	120
	AAATGTGAA TCTTTTAAAC ATCTTGATAA TTTGTTGTG AGAGCTGTTT ATTCTAAAAT	180
45	GTAATGAAAT TCAGTCTAGT TCTGCTGATA AAGATCATCA GTTTTGAAAG GTTACTGATT	240
	TTCTCTTCC CTCTTAGTTT TTTACCCAAT ATATGGAGAA GAGTAATGGT CAATCTTAAC	300
	ATTTTGTTTT AATGTTTTAA TAAAGCTGCT GGGCAGTGGT GCAGCAITCC TACCTAGTGT	360
50	CATAAAAGCA AAATACTTAC ATAGCTTTCT TAAAATATAG GAATGACATT ACATTTTAG	420
	GAGAAAGTAA GTTGCTTTGC ACCGCCTACT TAATTCCTTT CCATATATTG TGATACAAAC	480
55	TTTGAATAT GGAATCTTAC TATTGAATA GAAATGTGTA TGTATAATAT ACATACATAC	540
	ATAAGCATAT ATGTGTGTGT GTGTGTGTAT ATATATATAT ATGCATGCTG TGAAACTTGA	600
	CTACACAACA TAAATCACTT TTTAAATTC AGGAACGGGT AGTCTGACAC GGTGATTATC	660
60	CTTTGAGGC TGAATCCGTT ATTAACITGT TATTTAGGTT TTAATCCCAG TAGCAAGGGA	720

	TTCTAAGTTA GTTGCACTTA CATGATTATT GTGATTTAAA ACTAAGAATA AAGGCTGCAT	780
5	TTTCAAAGAT AAATTGGAAT TGCTGTTGGT GAAATAACAA CCAAAATACT GAATCTGATG	840
	TACATACAGG TTTCTACAGG AAGAGATGGT ATAATTTACA ATTTGGAGAT TTAATAACCA	900
	GGGCTACCCA GAAAAAGTGA CTTGATAACA TGGTACCAAT AAGTAAGGGA TGCTCTCTCG	960
10	GTTTGCTTTT GCCACTTTCA AGATTTTAAC TTCTCAGGTT ATTAATCAAA ATTATTGTAT	1020
	AAGTTAGCCA ATAGAATTTT TAGGTTAAAA CAACAGATGG GGGGTTTGTG GAGTGTTTAA	1080
15	TGTCATGGGC ATTTTITAGTA GCATAGACCC TTTGTTCTGC ATTTGAATGT TTCGTATATT	1140
	TTTGTTTCAC AGTTAATCTT CCTCCCCAA GTTTGCTATT CAAATCAACT GCCTGAATGA	1200
	CATTTCTAGT AGTCTGATGT ATTTTCTGA GGAATAGTTT GTGATTCCAA TGCAGGTGTC	1260
20	TTCATTACCA TTACCTCTAC ACTGCAGAAG AAGCAAACT CCTTTATTAG AATTACTGCA	1320
	CATGTGTATG GGGAAAATAG TTCTGAAAGG CTAGAATGAT ACAAGTGAGC AAAAGTTGGT	1380
25	CAGCTTGGCT ATGGAGTGGT GGCAATAATC TCTAAACATT CCAAAAGACC ATGAGCTGAA	1440
	CCTAAACTCC CTTGGGAATC TGAACAAAG GAATATGAAA ATTGCCATT T GAAAACGAC	1500
	CAGCTAATCT GGACCTCAGA GATAGATCAG CCAGTGGCCC AAAGCCATT CAAGTACAGA	1560
30	AATTATAGAG ACTACAGCTA AATAAATTG AACATTAAAT ATAATTTTAC CACTTTTGT	1620
	CTTTATAAGC ATATTTGTAA ACTCAGAACT GAGCAGAAGT GACTTTACTT TCTCAAGTTT	1680
35	GATACTGAGT TGACTGTCC CTTATCCCTC ACCCTTCCCC TTCCCTTTCC TAAGGCAATA	1740
	GTGCACAACT TAGGTTATTT TTGCTTCCGA ATTTGAATGA AAAACTTAAT GCCATGGATT	1800
	TTTTTCTTTT GCAAGACACC TGTTTATCAT CTTGTTTAAA TGTAAATGTC CCTTATGCT	1860
40	TTTGAAATAA ATTTCCTTTT GTAAAAAAA AAAAAAAA AAAAAA	1906

45 (2) INFORMATION FOR SEQ ID NO: 138:

(i) SEQUENCE CHARACTERISTICS:

- 50 (A) LENGTH: 1935 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 138:

55	TCTGAACATA TGCTAACAGA TCCCCCTGAG GGATTCTTGA TGGGCTGAGC AGCTGGCTGG	60
	AGCTAGTACT GACTGACATT CATTGTGATG AGGGCAGCTT TCTGGTACAG GATTCTAAGC	120
60	TCTATGTTTT ATATACATTT TCATCTGTAC TTGCACCTCA CTTTACACAA GAGGAACTA	180



	TGCAAAGTTA GCTGGATCGC TCAAGGTCAC TTAGGTAAGT TGGCAAGTCC ATGCTTCCCA	240
	CTCAGCTCCT CAGGTCAGCA AGTCTACTTC TCTGCCTATT TTGTATACTC TCTTTAATAT	300
5	GTGCCTAGCT TTGGAAAGTC TAGAATGGGT CCTGGTGCY TTTTACTTT GAAGAAATCA	360
	GTTCCTGCCT CTTTTTGAA AAGAAAACAA AGTGCAATTG TTTTACTG GAAAGTTACC	420
10	CAATAGCATG AGGTGAACAG GACGTAGTIN AGGCCTTCCT GTAAACAGAA AATCATATCA	480
	AAACACTATC TTCCCATCTG TTTCTCAATG CCTGCTACTT CTGTAGATA TTTCAITTTCA	540
	GGAGAGCAGC AGTTAAACCC GTGGATTTTG TAGTTAGGAA CCTGGGRTCA AACCCCTCTC	600
15	CACTAATTGG CTATGTCTCT GGACAAGTIT TTTTTTTTT TTTTTTTTAA ACCCTTCTG	660
	AACTTTCCT TCTATGTCT ACCTCAAAGA ATTGTGTGA GGCTTGAGAT AATGCATTTG	720
20	TAAAGGGTCT GCCAGATAGG AAGATGCTAG TTATGGATTI ACAAGGTGT TAAGGCTGTA	780
	AGAGTCTAAA ACCTACAGTG AATCACAATG CATTTACCCC CACTGACTTG GACATAAGTG	840
	AAACTAGCC AGAAGTCTCT TTTCAAATT ACTTACAGGT TATTCAATAT AAAATTTTGT	900
25	TAATGGATAA TCTTATTTAT CTAAACTAAA GCTTCCTGTT TATACACACT CCTGTTATC	960
	TGGGATAAGA TAAATGACCA CAGTACCTTA ATTTCTAGGT GGGTGCCTGT GATGGTTCAT	1020
30	TGTAGGTAAG GACATTTTCT YTTTTTCAGC AGCTGTGTAG GTCCAGAGCC TCTGGGAGAG	1080
	GAGGGGGTA GCATGCACCC AGCAGGGGAC TGAAGTGGGA AACTCAAGGT TCTTTTTACT	1140
	GTGGGTAGT GAGTGCCTT TCTGTGATCG GTTTCCTAG GGATGTTGCT GTTCCCTCC	1200
35	TTGCTATTCG CAGCTACATA CAACGTGGCC AACCCAGTA GGCTGATCCT ATATATGATC	1260
	AGTCTGGTG CTGACTCTCA ATAGCCCCAC CCAAGCTGGC TATAGGTTTA CAGATACATT	1320
40	AATTAGGCAA CCTAAAATAT TGATGCTGGT GTTGGTGTGA CATAATGCTA TGGCCAGAAC	1380
	TGAAACTTAG AGTTATAATT CATGTATTAG GGTCTCCAG AGGGACAGAA TTAGTAGGAT	1440
	ATATGTATAT ATGAAAGGA GGTATTAGG GAGAACTGGC TCCCACAGTT AGAAGGCGAA	1500
45	GTGCACAAT AGGCCGTCTG CAAGCTGGGT TAGAGAGAAG CCAGTAGTGG CTCAGCCTGA	1560
	GTTCAAAAAC CTCAAAACTG GGAAGCTGA CAGTGCAGCC AGCCTTCAGT CTGTGGCCAA	1620
50	AGGCCAAGAG CCCCTGGCAA CCAACCCACT GGTGCAAGTC CTAGATTCCA AAGGCTGAAG	1680
	AACCTGGAGT CTGATGTCCA AGAGCAGGAA GAGTGAAGA AAGCCAGAAG ACTCAGCAA	1740
	CAAGGTAGAC AGTGCTACC ACCAYAGTGG CCATACCAA GAGGCTACCG ATTCTTCTCT	1800
55	GCTACCTGGA TCCCTGAAGT TGCCCTGGTC TCTGCACCTT CTAAACCTAG TTCTTAAGAG	1860
	CTTTCATTA CATGAGCTGT CTCAAAGCCC TCCAATWAAT TCTCAGTGTA AGYTTCAAAA	1920
60	AAAAAAAAA AAAAA	1935

## (2) INFORMATION FOR SEQ ID NO: 139:

5

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1446 base pairs

(B) TYPE: nucleic acid

10

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 139:

15	NGCCCCCTTG GCACAAGTCA GATGAAGCAC GTTCTGCCGG GGAGGCCCTC AMCTTCCAGA	60
	GAGGACAGAC ACAGATTTC TGCTGGGGGA GGGAGGAGTC CACGCATCCT GATGCTGCCT	120
	GGAAGCTTAT TTTCCCGTGG CCAGGATGCA TTTCTCTGAG TGGAAACAGG TTCTTGCAATG	180
20	TGGATGTGTG TTTCCCCAGG CAGACGGCCC CTCTTTTCCC AGCACTTCCC TGCCTCCCCC	240
	AGGCCTCAGG CCAGCACCCA GTTCTCTCTC ACATGGCAGG TGAGCACAGA CTTCTAGTTG	300
25	GCAGGAGCTG AGGAGGGTGA ACAAACCCCG AGGGAGGCCG GGCCCTTGCT CCCGAGTTGG	360
	GGGGAGGGGG TGTGGCAACG TGCCCCCGC AGAGGCCACG CATGTTTGAC CAAAGCCCTC	420
	ATTGTGGTCC GAGGACAGCC TTTTCCCCAG GCCTCARAGC ATTGCTCATC CGTGCCAAAC	480
30	TGGGTAGGTG GATTTGAGCG GAAAGACTCC CAAAATGTGC CAAGAAATTTC CCRGTCCCAG	540
	GCAGGCGAGG GGAAACTAAG GGCAAGCAGG ATACAGGGCG AGGGATGTGG CAGGTGAGGG	600
35	GGCTCCCGCC TGTGCCCTT CTCTCACCA TGTCTCCCC ACCCTGCTC AGTCTCTCCG	660
	TCCCTTCAT CTCCGTCECC CTCTTTGAAG CTGTCCCCAT CTCAGTGTC GACCAGCCTT	720
	CTCTCAKCT GACCACCTC CTCTGACCSA CGCCCCCTCC TTGTCTGAAA AAAGGAGCCT	780
40	TGAATGGTGG AGGGAGGCAG TGGGAGAAA GGTCTCACC GACAGGTGG GAGAATGAGG	840
	TCAGCGGTGC TGGGGAACAG ATGGAGGGGG CAGTGGGGAC AGGGCTTGGG CAGACACCAG	900
45	CAGGAATAAT TTGAAATGTG TGAGGTGACT CCCCAGAGG CTTGGGCTTG GGCATTGGG	960
	AAAAGAATGA TGTCTGGAAG GGCTTAAGG ACACAGTGA CGAGGGGAGA GTCCTCATCT	1020
	GCTGGCATTT TGTGGGGTGT TAGTGCCAAA CTTGAATAGG GGCTGGGGTG CTGTCTTCCA	1080
50	CTGACACCCA AATCCAGAAT CCTGGTCTT GAGTCCCCAG AACTTTGCCT CTTGACTGTC	1140
	CCTTCTCTTC CTACCTCCAT CCATGGAAAA TTAGTTATTT TCTGATCCTT TCCCCGCT	1200
55	GGTCTAGCTC CTCTCCAAAC AGCCATGCCC TCCAAATGCT AGAGACCTGG GCCCTGAACC	1260
	CTGTAGACAG ATGCCCTCAG AATTGGGGCA TGGGAGGGG GSTGGGGGAC CCCATGATTC	1320
	AGCCACGGAC TCCAATGCCC AGCTCCTCTC CCCAAAACAA TCCCGACAAT CCCTTATCCC	1380
60	TACCCCAACC CTTTGGCGCT CTGTACACAT TTTTAAACCT GGCAAAAGAT GAAGAGAATA	1440

TTGTAA

1446

5

(2) INFORMATION FOR SEQ ID NO: 140:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1109 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 140:

TTTTTTTTTT	TTTGATATGA AATTGCTTTT CTCCATTGCA GAAATAAGCT AGGGAAACAC	60
TAACCCAAAA	ACTTTCTGTA GAGCTGTTCC TTTGGAGGCA GCATCACTTA TTGGCAGTAA	120
AGACTCAGTA	TAAAAGCACC AGCATCCCTA CTTGGGTGAT GGGGATTAAT TTTATAGCAT	180
TCCATTTTCC	TAGTGCCACA TGTGAAATG GATTTTGATG ATCTTAATCT ATATTCTACC	240
CTTATAATAA	AAGATCAAAA GATATATCTC CTATGAACAG ATTGGAGATA GGAGATGAAA	300
AGTTGGGAGG	ATGTCTTTAT TCTAATGTGA GGGTAGGGAA AATGTGGATA ACATTACTGG	360
GGTGARGGAG	GCATTGTCTT TTAGTTGGAG TTCTCATTTT TATTCTCCAG TACTGACTTG	420
TGGGGAAAGC	ATACTTTTTC ACTGCCAGGT ACTGAATGCA GAGGCTCAGT GAAGTATATA	480
TGTGGGAAGT	GCATGCATTT CGTTTATTAG CAAACATAGC TGGATTAAGA CAAAGTTGTT	540
GGTTTGAAAA	GGGGTTAAAG CCTTAAGTGA ACAAATCTAG CTAACAGTGA ATGAACTAGG	600
TAATATAACT	TGCATATTTT TAATTTCCCT TGGTTAAAGG TCCCCCATAC TTCTCTGTTT	660
GGAGACATGA	GAAGTATGAT TACTTCAGTG TTAGTTTCTT TAATTTTTTT TTTCCTTAT	720
TTGTCCCTTG	TCACTTTGTT GCAAGCTAGA AATCTGTGGG TTATACATAG GGCAGCTCTT	780
TGTGAAAGTG	GTTTATTCCA CTGGAGAAAG GGGATTGAAA ATCAGTTAGA ACCAATGTAT	840
TTCTTGCCCC	ACGGAACACT ATTCCTATAA GATAGCTGAA AGAAGCTGCT GTGAGGAGCT	900
CAGCTCCAAA	CACAGGATCA GCACCTTGTA TAGGAATTCC CATGAATTAT GACTTCTCAT	960
TCTGTTTTAT	CAGAGTGCAT ATATGTCCTA CTTGAGGAAA AGTAAAACAG TCATTTACGA	1020
AAGAAAGTCA	ATCTGTATCC TAAGCATTTT AATAAAAAGT TAAAACAAAA AATTAAAAGG	1080
GACACTCGAG	GGGGGGCCCG AAACCAAT	1109

55

(2) INFORMATION FOR SEQ ID NO: 141:

60 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 497 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 141:

10 TAGGACTAAC TTAAATTCCTT TTATTCATCT TTTATTTATT AAAAAATTTT ATTTCTTTGA 60  
10 ATTTTCCTGT AATTTCTTA RGCTCTCTA TAAATGTA TATTCATGTG AACCATACCT 120  
CATTATCCTT AACATTTACT CTCAAAAGC TTTTATTTT TATTTTTTTG AAGGTAGTTT 180  
15 TTCTGTGTGT ACTCTGTAAC ATGATTTTGC TTTCAAATCA TTGTTGTGCC CCCATACAAA 240  
ATGCCTTTTA TTTTGAGGA TCGTGGACTT TTTAGTATGG CATGAGTGTG CTAAAAGCCA 300  
GATATCTTTC CACATTCACT GGTGGCTTTG ACACCTAGTT TTTAATCTCC CATCCTTACT 360  
20 TTAAACCCCTG ACAGTGCAGT CCTCAGTCAG GGCCAGGACC GGGCTGAGGC CCTTTGTGGA 420  
GATGCTGCAC CACCAGCAGA AGGCTGAGAC CTGGTTACCT GTACCTGTTT ACTTGTAAATA 480  
AAAAGAATTA TCTAAAA 497

25

(2) INFORMATION FOR SEQ ID NO: 142:

30

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 269 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 142:

40 ATGAGGCAGA GGCAAGCTGC CTGCCAACCC CCTCCCTCAA GGAATGCCCT TGCCCAGGAA 60  
TGCCCACCAC ACATACCTC TTCTTTTTT CTAGTCAAAC TCTGTATTAT TCCTTGGCTT 120  
GCCTCCCTCC TTCTCTCCC TCTCAACCTT TACTTCTGG TTTCTATTTC ATGGGATTG 180  
45 GGGTGAAGT TAAACTTACA ACAGTGCCGC CAACACCAAG TCTTGAGGA AAAAAATACA 240  
AAGAAATTTA ACAAAAAAAA AAAAAAAA 269

50

(2) INFORMATION FOR SEQ ID NO: 143:

55

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1269 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 143:

	TTGATTGACT ATGGTCTCTC CGGCTACCAG GAAGAGTCTG CCGAAGTGAA GGCCATGGAC	60
5	TTCATCACCT CCACAGCCAT CCTGCCCTG CTGTTGGCT GCCTGGGCGT CTTCGGCCTC	120
	TTCCGGCTGC TGCAGTGGT GCGCGGAAG GCCTACCTGC GGAATGCTGT GGTGGTGATC	180
	ACAGCGCCA CCTCAGGGCT GGGCAAAGAA TGTGCAAAAG TCTTCTATGC TCGGGTGCT	240
10	AAACTGGTGC TCTGTGGCG GAATGGTGG GCCCTAGAAG AGCTCATCAG AGAACTCACC	300
	GCTTCTCATG CCACCAAGGT GCAGACACAC AAGCCTTACT TGGTGACCTT CGACCTCACA	360
15	GACTCTGGG CCATAGTTGC AGCAGCAGCT GAGATCCTGC AGTGCTTTGG CTATGTCGAC	420
	ATACTTGTC ACAATGCTGG GATCAGCTAC CGTGGTACCA TCATGGACAC CACAGTGGAT	480
	GTGGACAAGA GGGTCATGGA GACAACTAC TTTGGCCAG TTGCTCTAAC GAAAGCACTC	540
20	CTGCCCTCCA TGATCAAGAG GAGGCAAGGC CACATTGTGC CCATCAGCAG CATCCAGGGC	600
	AAGATGAGCA TTCCTTTTCG ATCAGCATAT GCAGCCTCCA AGCACGCAAC CCAGGCTTTC	660
25	TTTGACTGTC TCGTGCCGA GATGGAACAG TATGAAATG AGGTGACCGT CATCAGCCCC	720
	GGCTACATCC ACACCAACCT CTCTGTAAAT GCCATCACCG CGGATGGATC TAGGTATGGA	780
	GTTATGGACA CCACCACAGC CCAGGGCCGA AGCCTGTGG AGGTGGCCCA GGATGTTCTT	840
30	GCTGCTGTGG GGAAGAAGAA GAAAGATGTG ATCCTGGCTG ACTTACTGCC TTCCTTGGCT	900
	GTTTATCTTC GAACTCTGGC TCTGGGCTC TTCTTCAGCC TCATGCCTCC AGGGCCAGAA	960
35	AAGAGCGGAA ATCCAAGAAC TCCTAGTACT CTGACCAGCC AGGGCCAGG CAGAGAAGCA	1020
	GCACCTTAG GCTTGCTTAC TCTACAAGG ACAGTTGCAT TTGTTGAGAC TTTAATGGAG	1080
	ATTTGTCTCA CAAGTGGGAA AGACTGAAGA AACACATCTC GTGCAGATCT GCTGGCAGAG	1140
40	GACAATCAAA AACGACAACA AGCTTCTTCC CAGGGTGAGG GGAAACACTT AAGGAATAAA	1200
	TATGGAGCTG GGGTTTAACTA CTAATAAACA TCTCAAACAG TAAAAAATAA	1260
45	AAAAAAAC	1269

(2) INFORMATION FOR SEQ ID NO: 144:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1944 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 144:

60	AAAAGGCAAA CTATAGGATA ACACAGAGCC CTTTTTGAAA ATAAATTGGC ATTGGAGTGT	60
----	---	----

	TTTACCCTCT AGCTGTTTTA CTTAGAATGT AACATATGCT GCCTACCCAC CTCAAAATGT	120
	CTGTACTGCA AGAGGGCCCT GGGCCTCTGC TTTCCATATT CACGTTTGGC CAGAGTTGTA	180
5	GTCCCAAAGA AGAGCATGGG TGGCAGATGG TAGGGAATTG AACTGGCCTG TGCAATGGGC	240
	ATGGAGCACA AGGGGTACACA GCATGCCTCC TGCCTTACCG TGGCAGTACG GAGACAGTCC	300
10	AGAACATGGT CTTCTTGCCA CGGGTGTGTG TTGTCTCTGG TGGTGTGCA TGTCTGTGGC	360
	TCACCTTTAT TCTTGAACT GAGGTTTACC TGGATCTGGC TACTGAGGCT AGAGCCCACA	420
	GCAGAATGGG GTTGGGCCTG TGGCCCCCAA ACTAGGGGGT GTGGGTTCAT CACAGTGTG	480
15	CCTTTTGTCT CCTAAAGATA GGGATCTACT TTTGAAGGA ATTGTTCCTC CCAAATAAAT	540
	TTGCTTTTACC TTGGTCTTTT CTTTGTGCCC AGTATTCAAG TGGTATAGCT CTGAGCAGGG	600
20	TCACATTGG CCAAACCTGA CACTGTCTTG CTGCATTCTC CTTTGGCAA CATCAGGGTC	660
	AGAATTTCAG ATAGCCCTTC CTAGGGCACT GGACTTTCTG GCATGGGGGC TGTGTTTGCA	720
	CAAGTTATTT TCATGTTACC TGGAGAGTGT CCAGAGGCTG CTCTGAGGCT GAGGTGTGTT	780
25	CCCCCTTGCC TGGTTCACG TGTCAGAGGG ATACCATCCT AGGGTCTGGG AATCCAAGGC	840
	CACGAGACTC CTTGGTTTGT GGTCCGAGAT CCTGTACTAA GGAGGGTCTG GCCAGAGGAA	900
30	CAGACCAGCT TTTGCACAAT GAAGCGCAAG GGAACAAGTG GTTGCCTGG TGTCTTACCT	960
	GTCTTGAACC TGGTCTGTG GGCCATTGAA AAGTTAGATC TGTGATCTCT GGGGTTTTTG	1020
	TGCTTTGTT CAATGCTTCC ACTCTAGGGC AGGCAGAGCA GTCTATACTC TCCCAAGCCT	1080
35	GCTTGACCTC CAAGTAGAGC TGATACAGAG ATCTGTGAAT ATTGTGATAG AAATTCTTTG	1140
	GTATTCATAC ATTTTCAGCTG CAAGTCAGCA ATTTCCAGG TACCATGTAA GCTATAAAAC	1200
40	AGTCATCTT AAAGACAGAG GATAGCTGTG ACTCATGGGA TCATGAGGTC CATGGCTGGT	1260
	TGCAGGTTC CTTTTCCTT CTCAGGTTT TGTCCTTCC TGTGTGTCC CCAGCAAGGG	1320
	AGAGACTGTG GGGTGGATTG GGAGAACAGA TTAGGAGTAT AGCAAATGAA CCCAGAATGG	1380
45	AACAGTGGG AGCTAACTGT GAATGAGGAG AGTACCTGCT GCAGGACCTG GAGGTCAGGT	1440
	GTGAATGCTG TATTGGCACA GGAATAAAT ATCTGGCGT CTGGAGCCTT CACCTCTCCG	1500
50	TCAAGTCTT CCTGTGATAC TGCCATGGCA CAGGATCTGA GTTGCAGCTC TGCACCCTAA	1560
	ATCACACCCT GGGCATGTG TGGGCTGCAG GGCTGCCAGG TTCTGTACTT GTGTCCAGCT	1620
	GTGGCCCTGG ATGCTGGAGC TGGAGGTTT TCTGTGCTCA GACTGTAGCC TGTAGCTCTT	1680
55	GGCCTGTGTA GAGCCCCCTC CTGTGCCCTC AGTGGCTGTC GTTTGTTAAC ATCATCAGGA	1740
	AGATGGGAAA GGTACGGCAG AATTTTCTG CCCTACAAAG GGTGAAGAG AAAGGACACA	1800
60	GTATTTTCAT GAAATTACCA TATATCTTTG TTTTCTTCA ACGAAAAAGT TAATTGAGGC	1860

AATGTCATCT GCTCAAAGTT GAGTGGTTTA TTCACAATAA ACTGTAAGTT TCTGATTATA 1920

AAAAAAAAAA AAAAAAAAAA AAAG 1944

5

(2) INFORMATION FOR SEQ ID NO: 145:

10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1021 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 145:

TCGACCCACG CGTCCGGGGT GCGCAACGGG GAGTTCCGGC TGGAGACCCG TGCTCTGGGC 60

20

CGGCGCCTTC ACCATGGCCT CGGCAGAGCT GGACTACACC ATCGAGATCC CGGATCAGCC 120

CTGTCTGGAGC CAGAAGAACA GCGCCAGCCC AGGTGGGAAG GAGGCAGAAA CTCGGCAGCC 180

25

TGTGGTGATT CTYTTGGGCT GGGGTGGCTG CAAGGACAAG AACCTTGCCA AGTACAGTGC 240

CATCTACCAC AAAAGGGGCT GCATCGTAAT CCGATACACA GCGCCGTGGC ACATGGTCTT 300

CTTCTCCGAG TCACTGGGTA TCCCTTCACT TCGTGTTTTG GCGCAGAAGC TGCTCGAGCT 360

30

GCTCTTTGAT TATGAGATTG AGAAGGAGCC CCTGCTCTTC CATGTCTTCA GCAACGGTGG 420

CGTCATGCTG TACCGCTACG TGCTGGAGCT CCTGCAGACC CGTCGCTTCT GCGCGCTGCG 480

35

TGTGGTGGGC ACCATCTTTG ACAGCGCTCC TGGTGACAGC AACCTGGTAG GGGCTCTGCG 540

GGCCCTGGCA GCCATCCTGG AGCGCCGGGC CGCCATGCTG CGCCTGTTCG TGCTGGTGGC 600

CTTTGCCCTG GTGGTCGTCC TGTTCACGT CCTGCTTGCT CCCATCAGAG CCNTCTTCCA 660

40

CACCCACTTC TATGACAGGC TACAGGAGCG GGGCTCTCGC TGGCCCGAGC TCTACCTCTA 720

CTCGAGGGCT GACGAAGTAG TCCTGGCCAG AGACATAGAA CGCATGGTGG AGGCACGCCT 780

45

GGCACGCCCG GTCCCTGGCG GTTCTGTGGA TTTCGTGTCA TCTGCACAG TCAGCCACCT 840

CCGTGACTAC CCTACTTACT ACACAAGCCT CTGTGTGAC TTCATGCGCA ACTGCGTCCG 900

CTGCTGAGGC CATTGCTCCA TCTCACCTCT GCTCCAGAAA TAAATGCCTG ACACCTCCCC 960

50

ACAAAAA AAAAATAA ACTCGAGGGG GGGCCCGTA CCAATTGCG CCTATAAAGG 1020

T 1021

55

(2) INFORMATION FOR SEQ ID NO: 146:

60

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1285 base pairs

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 146:

	GGCACGAGGA GGGCCACGGC AGCCATCGCG CTTTGCAGTT CGGTCTCCTG GTGTACGGCC	60
	AACGCCAAGT AGGGGATTGC GTTCCCTCCA GTCGCAGACC CTATCAGATT TGGATATGTC	120
10	CTTCATATTT GATTGGATTT ACAGTGGTTT CAGCAGTGTC CTACAGTTTT TAGGATTATA	180
	TAAGAAAAC TGGTAACTGG TATTTCTTGG ATTGGATAAT GCAGGAAAA CAACATTGCT	240
15	ACACATGCTA AAAGATGACA GACTTGGACA ACATGTCCCA ACATTACATC CCACTTCCGA	300
	AGAACTGACC ATTGCTGGCA TGACGTTTAC AACTTTTGAT CTGGGTGGAC ATGTTCAAGC	360
	TCCAAGAGTG TGGAAAACT ACCTTCCTGC TATCAATGGC ATTGTATTTC TGGTGGATG	420
20	TGCAGACCAC GAAAGGCTGT TAGAGTCAAA AGAAGAACTT GATTCACTAA TGACAGATGA	480
	AACCATTGCT AATGTGCCA TACTGATTCT TGGGAATAAG ATCGACAGAC CTGAAGCCAT	540
25	CAGTGAAGAG AGGTTGCGAG AGATGTTTGG TTTATATGGT CAGACAACAG GAAAGGGGAG	600
	TATATCTCTG AAAGAACTGA ATGCCCGACC CTTAGAAGTT TTCATGTGTA GTGTGCTCAA	660
	AAGACAAGT TACGGAGAAG GCTTCCGCTG GATGGCACAG TACATTGATT AACACAACT	720
30	CACATTGGTT CCAGTCTCA ACGTTCAGGC TTA CTCAGAG ATTTGATTGC TCAACATGCA	780
	TAACCTGAAT TCAATAGACT TTTGCTGGTT ATAAACAGA TGTTTTTTAG ATTATTAATA	840
35	TTAAATCAAC TTAATTTGAA TGAGAATTGA AAACGTATTC AAGTAAGTTT GAGTATCACA	900
	ATGTTAGCTT TCTAATTCOA TAAAAGTACT TGGTTTTTAC AGTTTATAAT CTGACATCAC	960
	CCCAGCGCCA TTTGTAAAGA GCAACTTTCC AGCAGTACAT TTGAAGCACT TTTTAACAAC	1020
40	ATGAACTAT AAACCATATT TAAAAGCTCA TCATGTTAAA TTTTATATGT ACTTTTCTGG	1080
	AACTAGTTTT TAAATTTTAG ATTATATGTC CACCTATCKT AAGTGTACAG TTAATAATTA	1140
45	GCTTATTCAA TGATTGCATG ATGCCTTACA GTTTTCAATA ACTTTTTTTC TTATGCAAAC	1200
	GTCATGCAAT AAAACAACT CTAATGTTTG GCAAAAAAAA AAAAAAAA NTCGAGGGGG	1260
50	GGCCCGTACC CAATTGCCCC TAAAG	1285

55 (2) INFORMATION FOR SEQ ID NO: 147:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1386 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
60 (D) TOPOLOGY: linear



## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 147:

5	GGCACCAGGT GCGCAGGGG TCAGTGGTTC TCTCGGGTCT CGGACAGGT GAGCACCTG	60
	ATGAAGGCCA CGTCTCTGAT GCGGCACCTG GCGGGGTGCA GGAGATCGTG GCGCGCCTCC	120
	GCAAGGGCGS CGGAGACCGG TTACAGGTGA TTTCTGATTT TRACATGACC TTGAGCAGGT	180
10	TTGCATATAA TGGAAAGCGA TGCCCTTCTT CTTACAATAT TCTGGATAAT AGCAAGATCA	240
	TCAGTGAGGA GTGTCGGAAA GAGCTCACAG CGCTCCTTCA CCACTATTAC CCAATTGAGA	300
15	TCGACCCACA CCGGACCGTC AAGGAGAAGC TACCTCATAT GGTGGAATGG TGGACCAAAG	360
	CGCACAACTC CCTATGTCAG CAGAAGATTC AGAAGTTTCA GATAGCCAG GTGGTTAGAG	420
	AGTCCAATGC AATGCTCAGG GAGGGATATA AGACCTTCTT CAACACACTC TACCATAACA	480
20	ACATTCCTCT TTTTCATCTTT TCTGCGGGCA TTGGTGATAT CCTGGAAGAA ATTATCCGAC	540
	AGATGAAAGT GTTCCACCCC AACATCCACA TCGTGTCTAA CTACATGGAT TTTAATGAAG	600
25	ATGGTTTCTT CCAGGGATTT AAGGGCCAGC TGATACACAC ATACAACAAG AACAGCTCTG	660
	TGTGTGAGAA CTSTGGTTAC TTCCAGCAAC TTGAGGGCAA AACCAATGTC ATCCTGCTGG	720
	GAGACTCTAT CGGGGACCTC ACCATGGCCG ATGGGGTTCC TGGTGTGCAG AACATTCTCA	780
30	AAATTGGCTT CCTGAATGAC AAGGTGGAGG AGCGGCGGGA NCGCTACATG GACTCCTATG	840
	ACATCGTGCT GGAGAAGGAC GAGACTCTGG ATGTGGTCAA CGGGCTACTG CAGCACATCC	900
35	TGTGCCAGGG GGTCCAGCTG GAGATGCAAG GCCCCTGAAG GCGCAGGCTN CCAGNCCGCC	960
	TGCAGGCCGT GGTGAGGAGG GCGGCCTCCC CAGAGTCTGC TCCCCGTGA ACACAGAGCA	1020
	GANGCCAGGG TGGCCAGCAG TGGCTGGGTC CTTCCGCGCC CTTCCGTCTT CTTTCCCTG	1080
40	AGCACCTTCA TCACCAGAGG CTTGAAGGAA CCCCCTCATG TGGCAGGGCA CAGGCACTGT	1140
	TCCTGGTGAA CCTTGACCA CAGCATGTCA GTGCTCTAGG GATTGTCTAC TCCAGGGATT	1200
45	TTCTTCAAAA TTTTAAACA TGGGAAGTTC AAACAAATAT AATGTGTGAA ACAGATCAAA	1260
	ATTTTAAAAA TGAATAAAAA GCTGCTCTGA TTCAGGGGAT GTGGGTGCGG GTAGAACCTG	1320
	GACCTCTTGG CTTGGGGGCA CATGGGATGC TTCTAGGAAC ACAGTTTGAG AACCACCAAA	1380
50	AAAAAA	1386

## 55 (2) INFORMATION FOR SEQ ID NO: 148:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2098 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double

60

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 148:

5	AGCCCTTCTC CCCGCGCTTG GGACTCTGAC ATCTTAAGGC TGCACGGTCG TGTCTTGTTC	60
	TGGGTGAGGC CATGTCTGTG ATCCAAGGTT CCTGGAACTG ACACAGGAAG GGGCTGTGAA	120
	CCCTAAGTGG GTGTATCTC CTCCRACCGA GGCTTCTMAC CCTGGAGATG GCAGTTACTC	180
10	CTGGCCATGG TTGCTGAGCA TGGGCAGACC AGTGGAGGCC ACCCTACTGT GTTATCTGCG	240
	CCTTCRATGA AGTGAGACCC TTGGGGAGAA CGGGCTGTGG ATGAAGGAGT GGACTGCAGC	300
15	CTTGGCCTAG CCACTGGGCT GGGATCTTCT GGGTCATGTG ACTGTGTATC CAGGAGCAGA	360
	AACTGTATT CTCAGGATTC AGGATCTACC CAGCACCAAA GATGTATTTT CAGGAGAACA	420
	GACCTAGAAA TGGGCCTGTC TGGCATTTCA GAGTCAGGCA AAGCAGGCAG GGCCAGGGAG	480
20	CTTCTGTGGG TCTACACAAG AAGGTTCTTG TGAGGGCTAT CAGTGTGTGC CTCTAGCTT	540
	GCTGGTAACT TTGGCGCCTC CGCCAAGCCC TGCCAGACTC CCCTGGCTGT GATGGCATTC	600
25	TGTGCCATCC TGCCCTGTCC CCAGCCTCTG CAGGATGCCC TCCCTACCCA MCTYTYCCTG	660
	GGCCTTCCCT GTCCACTGGG CTGGATTCTT GTTCAAACCA CTGGACTGGC AGGGCAACGA	720
	CTTCTTCCCA CCTCAAGATG AGGTCTCTGC CCCCTGTCTT TGGCATAAAA ACACCTTTAA	780
30	AGCATGAGCC ATGTGCTTCT TTGCCCTTCT CTGTCTGTGT CCAATCTTCT GCCTCCCACT	840
	CACTCCCTGG GGACTATGGG ATCACTGTCC CCCCACCTGT GTGGCCACAC CATGTGTCTT	900
35	GTCAATCCAG AACTGCCTCT GAGCTCCAGG CTGACCACAG ATCAGCCACA GCCTGATGCC	960
	TGCAGCCCCA CTTTGCTCAC CCTTCCCCTC CCCTCTCTCT TCCTTCCACA CAGCAAGCCT	1020
	ACCTTTTYTCC ATCCATGCTC ACCATAGCCC CTTTCTTGT GACCTGGACC CTCCATGTGA	1080
40	CCTGGCTGAG ACTGTCAGCC TCCTGGAGGA GTGGGGTCCA CCTTCTTCTT GCCCTATGCA	1140
	GTGCAAGCTT CACTTCTCAC CCAGCAAGGT TGACTCATCT GCCTCCATGT CTCTGGGGCT	1200
45	TTGCTGTGTC CCTGAAACCT AGCTGGGCTG GTCTTGCTCC CAGCTTGCTT CCCCCTCTCT	1260
	GGATGTCCCT TTGCAGGCC CTGTCGTTCC TCCGGCACCA GTGTCTTGG CTGCCATGGC	1320
	AAGCTCATCA GGGGCTTGTG CCCTGGTCAC CAAGCATGGT AGCAGCTGCC TGCATTGTAT	1380
50	CTCCATCTGG TCACTGCAGG TGCCAACCCT TCATCCCCCA TGTPTTCTCT GGCCATGGAG	1440
	GGCTGACCTC CGTTCTCTGG GAATGTGGCT GAGCTGTGGT AACCAGCTAC ACCCCAGGTG	1500
55	CTCTTTCCAT GGTGGTGCTT GCTCATCTTG CTGATGCAAA CTAGGAAGTT AGGCTGCATC	1560
	TGGAGTGGC TTTCGCTGGA GAGGTGCTTT GCTGTCTCTC AGACTCAGTC ACTGTGTTCC	1620
60	CTCCCCGCTT CTCTATCTC CATGGCTGTT TGCAGCTCTC CCAGGTACTT TGGGGTCTGA	1680

GCTGGAATTC CTTTGTGGTT TGCTCTCTG CTTCTCACTC TTGTATTAAG AAGGATTCCA 1740  
 CAAAGGGAGA GTGGCATCCC TGCTGCTGCT GTGCCAGACC AGAGTTTCCT GAGGGGCCCT 1800  
 5 GACCCTAACC CTCCAGCTCA GCCCTGTACA CCTGACCCCTG TAAATGAGTG GGGTTTGCTG 1860  
 ACTGTAATCC CTGACACCAG TAAAACCAAA AGGACTCTTG GGGCTCAGT GTGAGAGCCA 1920  
 GGGTTACCTA CTCTGCCAAG TGAGGACAAA CTGCTAGGCT GTATCCCAT AATTTCAGGAT 1980  
 10 GAGAAACATT AACAATAAAA ATTTGTAGTA AACATAACCT CATGANGACT AAAAAAAAAA 2040  
 AAAAACYGG GGGGGGGCCC GTAACCCATT GGGCCCTTNG GGGGGNGTT TTAAAATT 2098

15

(2) INFORMATION FOR SEQ ID NO: 149:

20

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1847 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 149:

TCGACCCACG CGTCCGAACT GAGGCGGCGG CGGGAGCCCG TTGGKGTCTG GTCTTCGCGT 60  
 30 CGGCCCCGCG GACCAGACGC TGCCCCGGC GCGGGAGAA GATGGTGCK AGCGGCCTCG 120  
 GGGCCCCAC GCGCGCCAC GAGTGAGCCC AGCGGACCG CGGGCGTCCG CCGAGCAGCT 180  
 GGGCCGGCTG GGGCCGGGC GCGCANTGCC CGCCGGGCG GGTGGAGCT GATCAGAATA 240  
 35 ATGTTACGCA TCAACCCCTT GGAGAACCTG AAGGTGTACA TCAGCAGTCG GCCTCCCTTG 300  
 GTGGTCTTCA TGATCAGCGT AANGCCCATG GCCATAGCTT TCCTGACCCT GGGCTACTTC 360  
 40 TTCAAAATCA AGGAGATTAA ATCCCCAGAA ATGGCAGAG ATTGGAATAC TTTTCTGCTA 420  
 CGGTTCAATG ATTTGGACTT GTGTGTATCA GAGAATGAAA CCTCAAGCA TCTCACAAC 480  
 GACACCACAA CTCCGAAAG TACAATGACC AGCGGGCAGG CCCGAGCTTC CACCCAGTCC 540  
 45 CCCCAGGCC TGGAGGACTC GGGCCCGTG AATATCTCAG TCTCAATCAC CCTAACCTG 600  
 GACCCACTGA AACCTTTCG AGGGTATTCC CGCAACGTC CCCATCTGTA CTCAACCATC 660  
 50 TTAGGGCATC AGATTGACT TTCAGGCAGG GAAGCCCACG AGGAGATAAA CATCACCTTC 720  
 ACCCTGCCTA CAGCGTGGAG CTCAGATGAC TGCGCCCTCC ACGGTCAGTG TGAGCAGGTG 780  
 GTATTACAG CCTGCATGAC CCTCACGCC AGCCCTGGG TGTCCCGT CACTGTACAG 840  
 55 CCACCGCACT GTGTTCTGA CACGTACAGC AACGCCACG TCTGGTACAA GATCTTCACA 900  
 ACTGCCAGAG ATGCCAACAC AAAATACGCC CAAGATTACA ATCCTTTCTG GTGTTATAAG 960  
 60 GGGGCCATTG GAAAAGTCTA TCATGCTTTA AATCCAAGC TTACAGTGAT TGTTCAGAT 1020

	GATGACCGTT CATTAAATAAA TTTGCATCTC ATGCACACCA GTTACTTCCT CTTTGTGATG	1080
5	GTGATAACAA TGTMTTGCTA TGCTGTTATC AAGGGCAGAC CTAGCAAATT GCGTCAGAGC	1140
	AATCCTGAAT TTTGTCCCGA GAAGGTGGCT TTGGCTGAAG CCTAATTCCA CAGCTCCTTG	1200
	TTTTTTGAGA GAGACTGAGA GAACCATAAT CCTTGCCTGC TGAACCCAGC CTGGGCCTGG	1260
10	ATGCTCTGTG AATACATTAT CTTCGATGT TGGGTTATTC CAGCCAAAGA CATTTCAAGT	1320
	GCCTGTAACCT GATTGTGACA TATTTATAAA AATCTATTCA GAAATTGGTC CAATAATGCA	1380
15	CGTGCTTTGC CCTGGGTACA GCCAGAGCCC TTCAACCCCA CCTTGGACTT GAGGACCTAC	1440
	CTGATGGGAC GTTTCACGT GTCTCTAGAG AAGGATTCCT GGATCTAGCT GGTCAACGAC	1500
	ATGTTTTAC CAAGTCACA GGAGCATTCG GTCGCTGATG GGGTTGAAGT TTGGTTTGGT	1560
20	TCTTGTTC AATCAATATG TAGAGAACAT TTGAAACAGT CTGCACCTTT GATACGGTAT	1620
	TGCATTCCA AAGCCACCAA TCCATTTGT GGATTTTATG TGTCTGTGGC TTAATAATCA	1680
25	TAGTAACAAC AATAATACCT TTTCTCCAT TTGCTTGCA GGAAACATAC CTTAAGTTTT	1740
	TTTTGTTTTG TTTTGTFTT TTTGTTTTTT GTTTTCCTTT ATGAAGAAAA AATAAATAG	1800
30	TCACATTTTA ATACTACCAA AAAATGGACA AAAAAAGTCG AGGGGGG	1847

## (2) INFORMATION FOR SEQ ID NO: 150:

- 35 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1569 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - 40 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 150:

	GACGCTGACG AGAGAAGGCC TCTTCCTTGA GGGTTGGTGC TGTGTTGCAG TGACCGTGGC	60
45	GGATTACGCC AACTCGGATC CGGCGTCGT GAGGTCTGGA CGAGTCAAGA AAGCCGTAGC	120
	CAACGCTGTT CAGCAGGAAG TAAAATCTCT TTGTGGCTTG GAAGCCTCTC AGGTTCCTGC	180
50	AGAGGAAGCT CTTTCTGGG CTGGTGAGCC CTGTGACATC ATCGACAGCA GTGATGAGAT	240
	GGATGCCCAG GAGGAAAGCA TCCATGAGAG AACTGTCTCC AGAAAAAGA AAAGCAAGAG	300
	ACACAAAGAA GAACTGGACG GGGCTGGAGG AGAAGAGTAT CCCATGGATA TTTGGCTATT	360
55	GCTGGCCTCC TATATCCGTC CTGAGGACAT TGTGAATTTT TCCCTGATTT GTAAGAATGC	420
	CTGGACTGTC ACTTGCATG CTGCCTTTTG GACCAGGTTG TACCGAAGCA CTACAGCTG	480
60	GATGCTTCCC TGCCPTTTCG TCTGCGACCA GAGTCAATGG AGAAGCTGCG CTGTCTCCGG	540

	GCTTGTGTGA TCCGATCTCT GTACCATATG TATGAGCCAT TTGCTGCTCG AATCTCCAAG	600
	AATCCAGCCA TTCCAGAAAG CACCCCCAGC ACATTAAAGA ATTCCAAATG CTTACTTTTC	660
5	TGGTGCAGAA AGATTGTTGG GAACAGACAG GAACCAATGT GGAATTCAA CTTCAAGTTC	720
	AAAAACAGT CCCCTAGGTT AAAGAGCAAG TGTACAGGAG GATTGCAGCC TCCCGTTCAG	780
10	TACGAAGATG TTCATACCAA TCCAGACCAG GACTGCTGCC TACTGCAGGT CACCACCTC	840
	AATTTTCATCT TTATTCGAT TGTTCATGGA ATGATATTTA CTCTGTTTAC TATCAATGTG	900
	AGCAGGACA TCGGCATCA TCGAGTGAGA CTGGTGTTC AAGATTCCCC TGTCCATGGT	960
15	GGTCGAAAC TCGCAGTGA ACAGGGTGTG CAAGTCATCC TGGACCCAGT GCACAGCGTT	1020
	CGGCTCTTTG ACTGGTGGCA TCCTCAGTAC CCATTCTCCC TGAGAGCGTA GTTACTGCTT	1080
20	CCCATCCCTT GGGGGCAGCC TCGAGTGTAG TCCATTAGTA ATCAGATTCC AGTTTGGACA	1140
	GGTGGCTGG ATTGTATATC TCGTTAGTAA TGTACATGCT CTTCAGGTTT TAGGGCTCCT	1200
	GTTAGGGGAG GGAGAAATGT TGAATCAAGA GGGAAAACAA CTACTATGAT TTATAAACAT	1260
25	ATTTTAATGT AAAAATTTGC ATTTAAAGG AGTGGCCCTG TTTTCTGTGT TAAACCCCA	1320
	TTTGGTGCTA TTGAGTTTGT TCTTTATTCT TTTATCCCAG TGAAAATTGT TGATCTTGCT	1380
30	GTAGGAAAA ATTAACTCT TTGAATCTCC AAACAAGGAA GTTTCAGCAT TCCCTTATGG	1440
	ATCAGAGGAA CCTTAGAGGC CTGAAATGT TGCTTCCAGT TTAGCTGCCC CTCAAATTCA	1500
	AGTGAATATT TTCCCTTCTC CCTTTACCT TCTCCAGAAA TAAAGCAGGT GACAGGGTTT	1560
35	CAGAATCTT	1569

40 (2) INFORMATION FOR SEQ ID NO: 151:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1540 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 151:

50	CCCACGGTC CGGAAGGATT GACCAGTTAA CCAACATCTT AGCCCCCATG GCTGTTGGCC	60
	AGATTATGAC ATTTGGCTCC CCAGTCATCG GCTGTGGCTT TATTTGGGA TGGAACTTGG	120
	TATCCATGTG CGTGGAGTAC GTCCTGCTCT GGAAGGTTTA CCAGAAAACC CCAGCTCTAG	180
55	CTGTGAAAGC TGGTCTTAAA GAAGAGGAAA CTGAATTGAA ACAGCTGAAT TTACACAAAG	240
	ATACTGAGCC AAAACCCCTG GAGGGAAGTC ATCTAATGGG TGTGAAAGAC TCTAACATCC	300
60	ATGAGCTTGA ACATGAGCAA GAGCCTACTT GTGCCTCCCA GATGGCTGAG CCCTTCCGTA	360

CCTTCCGAGA TGGATGGGTC TCCTACTACA ACCAGCCTGT GTTCTGGCT GGCATGGGTC 420  
 TTGCTTTCCT TTATATGACT GTCCTGGGCT TTGACTGCAT CACCACAGGG TACGCCTACA 480  
 5 CTCAGGGACT GAGTGGGTC CATCCTCAGT ATTTTGATGG GAGCATCAGC TATAACTGGA 540  
 ATAATGGGAA CTGTAGCTTT TACTTGGCTA CGTCGAAAAT GTGGTTTGGT TCGGCAGGTC 600  
 10 TGATCTCAGG ATTGGCACAG CTTTCTGTT TGATCTTGTG TGTGATCTCT GTATTTCATGC 660  
 CTGGAAGCCC CCTGGACTTG TCCGTTCTC CTTTGAAGA TATCCGATCA AGGTTTCATT 720  
 AAGGAGAGTC AATTACACCT ACCAAGATAC CTGAAATTAC AACTGAAATA TACATGTCTA 780  
 15 ATGGGTCTAA TTCTGCTAAT ATTGTCCCGG AGACAAGTCC TGAATCTGTG CCCATAATCT 840  
 CTGTCACTCT GCTGTTTGCA GGCCTCATG CTGCTAGAAT CGGTCTTGG TCCTTTGATT 900  
 20 TAACTGTGAC ACAGTTGCTG CAAGAAAATG TAATTGAATC TGAAAGAGGC ATTATAAATG 960  
 GTGTACAGAA CTCCATGAAC TATCTTCTTG ATCTTCTGCA TTTCATCATG GTCATCCTGG 1020  
 CTCCAAATCC TGAAGCTTTT GGCTTGCTCG TATTGATTTC AGTCTCCTTT GTGGCAATGG 1080  
 25 GCCACATTAT GTATTCCGA TTGCCCCAAA ATACTCTGGG AAACAAGCTC TTTGCTTGCG 1140  
 GTCCTGATGC AAAAGAAGTT AGGAAGGAAA ATCAAGCAAA TACATCTGTT GTTTGAGACA 1200  
 30 GTTTAACTGT TGCTATCCTG TTACTAGATT ATATAGAGCA CATGTGCTTA TTTTGTACTG 1260  
 CAGAATTCCA ATAAATGGCT GGGTGTTTTG CTCTGTTTTT ACCACAGCTG TGCCTTGAGA 1320  
 ACTAAAGCT GTTAGGAAA CCTAAGTCAG CAGAAATTAA CTGGATTAAAT TTCCCTTATG 1380  
 35 TTAGGGGCCA TGGRAAAAAA ATTGGGAAAA GGAAAACTC AGTTTTAAAT ACGGGAGACT 1440  
 ATAATGGATA ACACTGRATT CCCCTATTTC TCATGAGTAG ATACAATCTT ACGTAAAAGA 1500  
 40 GTGGTATGTC ACGTGAATTC AGTTATCATT TGACAGATTC 1540

45 (2) INFORMATION FOR SEQ ID NO: 152:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1719 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 152:

55 TACTTATGAG GTCAATTGGA AATAAGAACA CCATTTTACT GGGTCTAGGA TTTCAAATAT 60  
 TACAGTTGGC ATGGTATGGC TTTGGTTCAG AACCTTGGAT GATGTGGGCT GCTGGGGCAG 120  
 60 TAGCAGCCAT GTCTAGCATC ACCTTTCCTG CTGTCACTGC ACTTGTTCGA CGAACTGCTG 180

	ATGCTGATCA ACAGGGTGTG GTTCAAGGAA TGATAACAGG AATTCGAGGA TTATGCAATG	240
	GTCTGGGACC GGCCCTCTAT GGATTCAATTT TCTACATATT CCATGTGGAA CTTAAAGAAC	300
5	TGCCAATAAC AGGAACAGAC TTGGGAACAA ACACAAGCCC TCAGCACCAC TTGTAACAGA	360
	ATTCCATCAT CCCTGGCCCT CCCTTCCTAT TTGGAGCCTG TTCAGTACTG CTGGCTCTGC	420
10	TTGTTGCCTT GTTTATTCCG GAACATACCA ATTTAAGCTT AAGGTCCAGC AGTTGGAGAA	480
	AGCACTGTGG CAGTCACAGC CATCCTCATA ATACACAAGC GCCAGGAGAG GCCAAAGAAC	540
	CTTTACTCCA GGACACAAAT GTGTGACGAC TGAAATCAGG AAGATTTTTC TATCAGCACC	600
15	CAGGTCTTAG TTTTCACCTC TAGTCTGGA TGTACATTCC ATTTCCATCC ACAGTGTACT	660
	TTAAGATTGT CTTAAGAAAT GTATCTGCAT GAACTCCGTG GGAATAAAG GAAGTGGGAA	720
20	CTTAGAACCA GACAGTTTTC CAAAGATGTT ACAATTCTTT TTGAAAAACC TTTTGTATTAT	780
	TAGCACC AAT TTCTYGCCAC TAAGCTATTT GTTTTATTAT ACATCCTTTA ATTAAAACT	840
	ATATATGTAA CTCTTAGAT ATTAGCAAAT GTCTCTGCTA CCATTTCTTT AAGGTGTGA	900
25	GCTTTAACTC TATGCTGACT CAGTGAGACA CAGTAGGTAG TATGGTTGTG GACCTATTTG	960
	TTTTAACATT GTAAAAATTT GAGTCAGATT TTAATATTGT AAAATCTTGG GTCAAATAAT	1020
30	TCAAAGCCTT AATGCAGATG CACTAAAACA AAGAAATGGT AAATGAATG TTTGCATTTA	1080
	AAAAAAAAA CTCTTAAGAA AACTGTACTA AATCTGAATC ATGTTTGTAG CTGTGTTGCA	1140
	GTACTTTTAA ACATTATTCA CTACTGTTTT TGAAGTGAGA AAGTATCAGC CATTTAGCAT	1200
35	TTAAGTTGGG GTATTTAGAG CCTGTAATCT AAATGCTGGC TCAAATTTAT TCCCCAGCTA	1260
	CTCTTTATAC CACTATTCTT TTAATGTTTG CATAATCATA AGCACCTCAA CACTTGAATA	1320
40	CATAATCTAA AAATTATATA GTAAAGCTGG TAGCCTTGAA AATGTCAGTG TGATATCTAT	1380
	TATGTAGATA AATATATATA GTGGCCTTTC AGGACTGTCA CAGTAACACT TTATTTACAG	1440
	AGCTAATGTT TGTCTTAAAT TTTCAGGACC CTAGAGGAGA GCTTTATACA ATTACCGATG	1500
45	TGAATTTCTC TAAAGTGTAT ATTTTGTGT CAGTTATAT TATTTAAAAA AGTGTACTT	1560
	TGTAAAAATT GTATATAAAG AACTGTATAG TTTACACTGT TTTTCATCTG TGTGTGGTTA	1620
50	TGCTTAATG CTTTTTAAAC TTGGAACACT CACTATGGTT AAATAAGGTC TTAAAGAAA	1680
	TGTAAATATT YTGTTAATAA AGTTAAATAT TTTAATGAT	1719

55

(2) INFORMATION FOR SEQ ID NO: 153:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 863 base pairs

(B) TYPE: nucleic acid

60

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 153:

5  
GGCAGGAGGG AAGCCGGGAC GATGTCGCA TGACAACCGA CGTTGGAGTT TGGAGGTGCT 60  
TGCCTTAGAG CAAGGGAAC AGCTCTCATT CAAAGGAAC AGAAGCCTCT CCCTCAGTGG 120  
10 TAGGAGACA GCCAGGAGCG GTTTCTGGG AACTGTGGGA TGTGCCCTTG GGGCCCGAG 180  
AAACAGAAG GAAGATGCTC CAGACCAGTA ACTACAGCCT GGTGCTCTCT CTGCAGTTCC 240  
TGCTGCTGTC CTATGACCTC TTTGTCAATT CCTTCTCAGA ACTGCTCCAA AAGACTCCTG 300  
15 TCATCCAGCT TGTGCTCTTC ATCATCCAGG ATATTGCAGT CCTCTTCAAC ATCATCATCA 360  
TTTCTCAT CAT GTTCTTCAAC ACCTTCGTCT TCCAGGCTGG CCTGGTCAAC CTCCTATGCC 420  
20 ATAAGTTCAA AGGGACCATC ATCCTGACAG CTGTGTACTT TGCCCTCAGC ATCTCCCTTC 480  
ATGTCTGGGT CATGAACCTA CGCTGGAAAA ACTCCAACAG CTTCATATGG ACAGATGGAC 540  
TTCAAATGCT GTTTGTATTC CAGAGACTAG CAGCAGTGT GTACTGCTAC TTCTATAAAC 600  
25 GGACAGCCGT AAGACTAGGC GATCCTCACT TCTACCAGGA CTCTTTGTGG CTGCGCAAGG 660  
AGTTCATGCA AGTTCGAAGG TGACCTCTTG TCACACTGAT GGATACTTTT CCTTCCTGGA 720  
30 TAGRAGGCCA CATTTGCTGC TTTGCAGGGG AGAGTTGGGC CCTATGCATG GGGCAAAACA 780  
GGTGGGATTT TCCAAGGGAA GGGTTCAGAA TTAGGCNTGT TGTTCAGCC ATTTCCAAGG 840  
AAGGGGAAGG GTTCCCTNC CCT 863  
35

(2) INFORMATION FOR SEQ ID NO: 154:

40 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1101 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
45 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 154:

50 AACAGCAAAA AAGAATGATT TCTTCTGAAA TTGTGGAACA TGAGGATTCA AGTTTTTATT 60  
TTGTTACTAG GTGCTGGAGG AACATCCAG TTCAACAAGC CCCCATCTCT TCCTCTGGAG 120  
CCAGAGCCTG CGGTGAATC AAGTCCAAC GAAACATCAG AACAAATAAG AGAGAAATAA 180  
55 GAATAGAATG AATGACCCCA AAATARGGTT TTCTTGGGCG AGGATGTGCT GGATTAGGAA 240  
AGGTGACATG ACACAGGCAG AGCAGAGTGG CACCCACCAC AGAATACAGT GTGTGTTATT 300  
ACGAGGAGCC AGCAGTTGAG CCTAAGGTCC TTCTACCTAC CTGGTATTGG CATTTGAGGT 360  
60



	CGGAAACCCCT CTAAGCCCC ATAAGCCAGG AAAAGTGAAA AGAGAACACA GTTCCTTTAA	420
	GAAGTGGCAG CAAGGCTTGA GGCCTTATGT ATGTAGCTGA GTCAGCAAGG TACATGATGC	480
5	TGTCCTGCTTT CAAAAGGACT TTTCTCTCCT AGCTGACTGA CTCTTTCCTT AGTTCAAGGA	540
	ACAGCTGAGA CAGACCTCTG CTGAGTAGCT CTGTGATGAC AAAGCCTTGG TTAACTGAG	600
10	GTGATCCTCA GGTGTGAGG TTTATTAGTC CCCAAGGCAA ACACAAATAT TAGATTAATA	660
	ATCCAACCTTT AATAGTATAC ATTTAAAAGA AAAAAACAA AAGCCCTGGA AGNITGAGGC	720
	CAAGCTGCT GAGTATTGCA GCTGCATTG CCCAAAGGGA ATCCAGAACA AGTCCCTCCC	780
15	TGTATTTTGT TCTTGAGAGG GGTGAGTCTA GAAGCTAGAT CCTATCAGGA TGAGGAGCAG	840
	CAGCCCAGGG CTGTCTGGA TCAGCACCAA CGATTTTAAA GAAAAAGGA AGAGTTTCTT	900
20	AGATGAGTAA TTGTTATTGA AGATAGTCAG TGATAACCAC TGACCAGATG CTATCAATAC	960
	ACTATGTGTC CTTTITAGAA TAAAGATTAC ATATCATCAT TCCTTTGGGG AAAATTGTTA	1020
	TTGAGGTATA AAAACAAGAG ATTATAATAA AAAANTAAAA GAACCCCTAAA AAAAAAAAC	1080
25	CTCGTGCCGA ATCCCTGCA G	1101
30	(2) INFORMATION FOR SEQ ID NO: 155:	
	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 2031 base pairs	
	(B) TYPE: nucleic acid	
35	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 155:	
40	CAATTAACCC GTTTGAGGCC TAGGTTGTTT GGCAAGCCCC NGGCTAAAG TTTTAATTG	60
	GCAGAGCCAA GGCCTGAAA GGAAGGGAAA GGGGAGGTA GCGGGAGGT AGCAGGTGAG	120
45	TTCCTAGGGC TGAAGGTTT AGCAGCAGCC TGGTGAGTG CCCTGTCATC AAGACAAACC	180
	CACGGTCTC CTGGGTGCTT ACCAAGCTTG GTTTGTACAA AAGCAAGGTG GGAGTCTATT	240
	TTTGTACATG AGATACATCA CACTTACCTG TGGGCCAGTA TTGTGAAGTG AGTCTGAGTT	300
50	GTTTACACTG ATGCCTTCCC TGCCACCAC AAATTGTGTA CATAGTCTTC AGAATGATAC	360
	CACCCCTTTC CCCAGTCCC AACCAGAGC TGGTCTAGG CCTGTGTTAT ATGTCATATT	420
55	TAGCGTTTTT ATATATGACC TTTGATTCTT GTTGTGTA TTTTAGCACA GTGTATGCAC	480
	CTTCATTTAA ATACATCTGT GTGCATACAG ATACGCATAT ATGTGTGTC GTATGCATAT	540
	ATCTCTCATC TGTAGTTTCC AAGAGTTCAG CTGAAGCAGA TGGAGTCTG CAGCCCAGGA	600
60	GACACCTGC ATCCCTGCTA ATAGTGTGTA CCACAAGTAT TAGTGAGTCT TCCTTATTAA	660

	TATTTTCATT TCAGAAGACT GAAGCAAAGC TGATAGTGT TGCTGTTTCT TTGGCAGCTA	720
5	AGTGAGGGTC TTGGGATGAC TTGCTGTGTT CCTCAAGCTG CACTTTGGGG CCATCTCTGC	780
	AGTATTAAGC CCCCTTTTGT CTGGTGGTA CTCTGTCTGT GCCTGTGTGT GTGTGTGATA	840
	GTCACCTCTG CATGGCTTCC ATGTCTGGTT TGTGGCATT GGGGATAAGT GCTGAACCAG	900
10	AGCATTGCA GTTGTGTTGA GGCCTCGTTG CCAATGATAG ATCACTCCTG TTGACCTGGT	960
	ATGTCTGCTT GCTGTCTGCT TTTCCTTGCT TTCTCTTGA AGAGGAAAGG ACTCTGGTCA	1020
15	GGCCCAGGCT GAGTGAGATG AGCTGCAGCT GGCTCATGGC CTTCTTAGAG CAGAGAGAGG	1080
	AGTATGTCAT TTTACTAAGT TCCTAAACAA ACATTTATGC AGGCAACACT CCTTGAGAT	1140
	CCAGAACTG AGGCACAATA GGGTTATGAC TTGCTCAAGA ATATGTAGCT GCTAGGGGGT	1200
20	AAATCAAGGC ATCACAATTT CTGTTTCAGG GGCAGGAATA GGCTGTGAAT TGCTAGCACT	1260
	TTTTTTTAA GCAATTACTT TTTGACTTGT TCCTCTGAAA GTGCAAGAGG CGTACACCTT	1320
25	TCCCAAATGT AGACTAGAAT CTGCAGGATG CCACCCACTG TATAGTTCTG CTTTCCCAGA	1380
	GAGGAAGAAC TTTTAGAAAC CAAATGATCT TAATTGTTAT TGCCCAACCC TGGCTTTTCC	1440
	GGGTAGAAAA TTCACAGTAG GAATGATTGT TAAGAGAGAG TGCTTGAAC CATGGGTTAA	1500
30	CAGGAAAGGC TACCTAACTT CACATATCTG CAACCAGAGC AGCCACCAAG CATTACTTAG	1560
	CAGCAGGAAA ATGATTGTAT TTGAGTTCTT GTGTGTCCAA AACTGAGGCA CCATGTTCTT	1620
35	TGAAAACATG CCACCTCAAG GCTGGGCGCG GTGGCTCACA CCTGTTAATC CCAGCACTTT	1680
	GGGAGGCCGA GCGGGGCGGA TCACCGGAGT CGGGGAGTTT GAGACCAGCC TGGACCAACA	1740
	TGGGAGAAAC CCCATCTCTA CCTAAAAATA CAAAATTAGC CGGGCGTGGT GGCATGCGCC	1800
40	TATAATCTCA GCTACTTGGG AGGGYTGAGG CAGGRGAATT GCTTGAACCC RGGANGCGG	1860
	AGGTTTGGCG TTGAGTTGAG GATCGTGCCA TTGCACTTCC GGGCCTTGGG GCAACAACAG	1920
45	CAAAAAYTCC GTCTTCAAMW MRTGCCGAAT TCGATATCAA GCTTATCGAT ACCGTCGACC	1980
	TCGAGGGGGG GCCCGGTACC CAATTCGCCC TATAGNGATC GTATTACAAT C	2031

50

(2) INFORMATION FOR SEQ ID NO: 156:

(i) SEQUENCE CHARACTERISTICS:

55

- (A) LENGTH: 1981 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 156:

	CCTGCACCCCT GAGCCCTTCA CCCCTCCGAG TTCCCCCAG GTTGGCTTCC TTCGATTCCCT	60
	TTTCTTGTA TCAACGTTTG ATTGGAAGAA CAACCCCTC TTTGTCAACC TCAATAATGA	120
5	GCTCACTGTG GAGGAGCAGC TCGGGCACAG CTCMCCGTYA TGGTCATTGT TACCCCCCAA	180
	GACCGCAAAA ACTCTGTGTG GACACAGGAT GGACCCCTCAG CCCAGATCCT GCAGCAGCTT	240
10	GTGGTCCCTGG CAGCTGAAGC CCTGCCCATG TTAGAGAAGC AGCTCATGGA TCCCCGGGGA	300
	CCTGGGGACA TCAGGACAGT GTTCCGGCCG CCCTTGGACA TTTACGACGT GCTGATTCCG	360
	CTGTYTCTC GCCATATCCC GCGGCACCGC AGGCTTGTGG ACTGCCAGY TGCTCTCTC	420
15	TGCCGGGGCC TGCTCAGCCA GCCGGGGCCC TCATCCCTGA TGCCCGTGCT GGGTATGAT	480
	CCTNCTCAGC TCTATCTGAC GCAGCTCAGG GAGGCCCTTG GGGATCTGGC CCTTTCTCTC	540
20	TATGACCAGC ATGGTGGAGA GGTGATTGGT GTCCTCTGGA AGCCACCAG CTTCAGCCG	600
	CAGCCCTTCA AGGCCTCCAG CACAAAGGGG CGCATGGTGA TGTCTCGAGG TGGGGAGCTA	660
	GTAATGGTGC CCAATGTTGA AGCAATCTG GAGGACTTTG CTGTGCTGGG TGAAGGCCCTG	720
25	GTGCAGACTG TGGAGGCCCG AAGTGAGAGG TGGACTGTGT GATCCCAGCT CTGGAGCAAG	780
	CTGTAGACGG ACAGCAGGAC ATTGGACCTC TAGAGCAAGA TGTCAGTAGG ATGACCTCCA	840
30	CCCTCCTTGG ACATGAATCC TCCATGGAGG GCCTGCTGGC TGAACATGCT GAATCATCTC	900
	CAACAAAACC CAGCCCCAAC TTTCTCTCTG ATGCTCCAGC ATTGGGGCAG GGGCATGGTG	960
	GCCCATGTAG TCTCTGGGC CTCACCATCC CAGAAGAGGA GTGGGAGCCA GCTCAGAGAA	1020
35	GGAAGTGAAC CCAGGAGATC CATCCACCTA TTAGCCCTGG GCCTGGACCT CCCTGCGATT	1080
	TCCCACCTCT TCTTAGTCT TCTTCCAGAA ACAGAGAAGG GGATGTGTGC CTGGGAGAGG	1140
40	CTCTGTCTCC TTCTGCTGC CAGGACCTGT GCCTAGACTT AGCATGCCCT TCACTGCAGT	1200
	GTCAGGCCCT TAGATGGGAC CCAGCGAAAA TGTGGCCCTT CTGAGTCACA TCACCGACAC	1260
	TGAGCAGTGG AAAGGGGCTA TATGTGTATG AATAGACCAC ATTGAAGGAG CACAATGCCC	1320
45	TCCTGTGTGT ATGCCACTTC CCAGGGTGA GACAGTGGAA AAGAACCGAG GACAGGAAAG	1380
	GATTGGGTAG GTGAAGGGGT CAGGGGACTG GTAGTCACCC AATCTTGGAG AGGTGCAAAA	1440
50	AGCACTGGGG GCTACCGGTT AGCTGCATCT GCCCTGGCTG TTTGCCCGTT CATGTCACAA	1500
	ACTGCCACTA CTATGTACCT GCAGTGGGTT TGCAGAGATG GGGGAGACTC AAGTCTTACT	1560
	CCCCAGGAGC TCCCAGGGCC CAAGGAGGAG AATGCTGCCT CCTTTCAGTC TGGTCTACAC	1620
55	CCACTTTCTG GTAGCCTCTC TGCTTCCTGT AATTCTGGCT GTTTTCCAG ACTCAGCTCA	1680
	AATAGTGCCC CTCCCTAAGC CCATCCCTCG CCCCCAGCCT GAGGTGATCT TTCCCTCCTC	1740
60	TGAAGTATTA GAGCAGTTAC TGCTGTGTCA GTTCGTTTGG CAGGCACACA CAGTGGCATA	1800

AATTCCTATTG TTTTGAACTC TGATTAAAA TTA AATTGCA GCTGGGCGTG GTGGCTCATG 1860  
CTGTGAATCC CAACACTTAG GGAGTMAGGR GAATCACTTG ASCYCAGGAG TYCTAGACCA 1920  
5 ATCTGGGCAA MAGAGAGACC CCATCTCTTT TAAATAAAAA GTTAAATTGC TTA AAAAAA 1980  
A 1981

10

(2) INFORMATION FOR SEQ ID NO: 157:

(i) SEQUENCE CHARACTERISTICS:  
15 (A) LENGTH: 915 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 157:

GAATTCGGCA CGAGCGCGGC CATGGCGCTC CTGCTTTCGG TGCTGCGTGT ACTGCTGGGC 60  
GGCTTCTTCG CGCTCGTGGG GTTGGCCAAG CTCTCGGAGG AGATCTCGGC TCCAGTTTCG 120  
25 GAGCGGATGA ATGCCCTGTT CGTCAGTTT GCTGAGGTGT TCCCGCTGAA GGTATTTGGC 180  
TACCAGCCAG ATCCCCCTGAA CTACCAATA GCTGTGGGCT TTCTGGAAGT GCTGGCTGGG 240  
30 TTGCTGCTGG TCATGGGCCC ACCGATGCTG CAAGAGATCA GTA AACTTGT CTTGATTCTG 300  
CTCATGATGG GGGCTATCTT CACCTTGGCA GCTCTGAAAG AGTCACTAAG CACCTGTATC 360  
CCAGCCATTG TCTGCCCTGG GTTCTCTGCTG CTGCTGAATG TCGGCCAGCT CTTAGCCCAG 420  
35 ACTAAGAAGG TGGTCAGACC CACTAGGAAG AAGACTCTAA GTACATTCAA GGAATCCTGG 480  
AAGTAGAGCA TCTCTGTCTC TTTATGCCAT GCAGCTGTCA CAGCAGGAAC ATGGTAGAAC 540  
40 ACAGAGTCTA TCATCTTGTT ACCAGTATAA TATCCAGGT CAGCCAGTGT TGAAAGAGAC 600  
ATTTTGCTA CCTGGCACTG CTTTCTCTTT TTAGCTTTAC TACTCTTTTG TGAGGAGTAC 660  
ATGTTATGCA TATTAACATT CCTCATGTCA TATGAAAATA CAAAATAAGC AGAAAAGAAA 720  
45 TTTAAATCAA CCAAAATTCT GATGCCCCAA ATAACCACTT TTAATGCCTT GGTGTAAGTA 780  
TACCTCTGAA CTTTTTCTG TGCCTTTAAA CAGATATATA TTTTTTTTWA ATGAAAATAA 840  
50 AACCATATAT CCTATTTTAT TTCTCTCTT TAAAACCTTA TAACTATAA MAAAAA 900  
AAAAA AAAAA CTCGA 915

55

(2) INFORMATION FOR SEQ ID NO: 158:

(i) SEQUENCE CHARACTERISTICS:  
60 (A) LENGTH: 2117 base pairs

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 158:	
	AGAGCGAAGC GAGGGTGGCG CGGGTCCGGG CATGAAGCTG GGCCGGGCCG TGCTGGGCCT	60
	GCTGCTGCTG GCGCCGTCCG TGGTGCAGGC GGTGGAGCCC ATCAGCCTGG GACTGGCCCT	120
10	GGCCGGCGTC CTCACCGGCT ACATCTACCC GCGTCTCTAC TGCTCTTCG CCGAGTGTG	180
	CGGCAGAAG CGGAGCCTTA GCCGGGAGGC ACTGCAGAAG GATCTGGACG ACAACCTCTT	240
15	TGGACAGCAT CTTGCAAAGA AAATCATCTT AAATGCCGTG TTTGGTTTCA TAAACAACCC	300
	AAAGCCCAAG AAACCTCTCA CGCTCTCCCT GCACGGGTGG ACAGGCACCG GCAAAAATTT	360
	CGTCAGCAAG ATCATCGCAG AGAATATTTA CGAGGGTGGT CTGAACAGTG ACTATGTCCA	420
20	CCTGTTTGTG GCCACATTGC ACTTTCCACA TGCTTCAAAC ATCACCTTGT ACAAGGATCA	480
	GTTACAGTTG TGGATTGAG GCAACGTGAG TGCTGTGCG AGGTCCATCT TCATATTTGA	540
25	TGAAATGGAT AAGATGCATG CAGGCCTCAT AGATGCCATC AAGCCTTTC TCGACTATTA	600
	TGACCTGGTG GATGGGTCT CCTACCAGAA AGCCATGTC ATATTCTCA GCAATGCTGG	660
	AGCAGAAAGG ATCAGAGATG TGGCTTTGGA TTTCTGGAGG AGTGGAAAGC AGAGGGAAGA	720
30	CATCAAGCTC AAAGACATTG AACACGCGTT GTCTGTGTCG GTTTTCAATA ACAAGAACAG	780
	TGGCTTCTGG CACAGCAGCT TAATTGACCG GAACCTCATT GATTATTTTG TTCCCTTCCT	840
35	CCCCCTGAA TACAAACACC TAAAAATGTG TATCCGAGTG GAAATGCAGT CCCGAGGCTA	900
	TGAAATGAT GAAGACATTG TAAGCAGAGT GGCTGAGGAG ATGACATTTT TCCCCAAGA	960
	GGAGAGAGTT TTCTCAGATA AAGGCTGCAA AACGGTGTTC ACCAAGTTAG ATTAATTA	1020
40	CGATGATTGA CAGTCATGAT TGGCAGCCG AGTCACTGCC TGGAGTTGGA AAAGAAACAA	1080
	CACTCAGTCC TTCCACACTT CCACCCCGAG CTCTTTTCCC TGAAGAGGA ATCCAGTGAA	1140
45	TGTTCCGTGT TGATGTGACA GGAATTCTCC CTGGCATTGT TTCCACCCCG TGGTGCCTGC	1200
	AGGCCACCCA GGGACCACGG GCGAGGACGT GAAGCCTCCC GAACACGCAC AGAAGGAAGG	1260
	AGCCAGCTCC CAGCCCACTC ATCGCAGGCG TCATGATTTT TTACAAATTA TGTTTTAATT	1320
50	CCAAGTGTGT CTGTTTCAAG GAAGGATGAA TAAGTTTAT TGAAAATGTG GTAACTTTAT	1380
	TTAAAATGAT TTTTAACATT ATGAGAGACT GCTCAGATTC TAAGTTGTG GCCTTGTGTG	1440
55	TGTGTTTTTT TTTAAGTTCT CATATTATT ACATAGACTG TGATGTATCT TTACTGGAAA	1500
	TGAGCCCAAG CACACATGCA TGGCATTGTG TCCACAGGAG GGCATCCCTG GGGATGTGGC	1560
60	TGGAGCATGA GCCAGCTCTG TCCCAGGATG GTCCAGCGG ATGCTGCCAG GGGCAKTGAA	1620

GTGTTTAGGT GAAGGACAAG TAGGTAAGAG GACGCCTTCA GGCACCACAG ATAAGCCTGA 1680  
 AACAGCCTCT CCAAGGGTTT TCACCTTAGC AACAAATGGA GCTGTGGGAG TGATTTTGGC 1740  
 5 CACACTGTCA ACATTTGTTA GAACCACTCT TTTGAAAGAA AAGTATTTC AACTTGTAC 1800  
 TTGCCAGTCA CTCGGTTTTC CAAAAGGTGG CCCTTCACTG TCCATTCCAA ATAGCCCACA 1860  
 CGTGCTCTCT GCTGGATTCT AAATTATGTG AATTTTGCCA TATTAAATCT TCCTCATTTA 1920  
 10 TACTATTATT TGTTACGTTT AATCAGAATC CCCGAAACCT CCTATAAAGC TTAGCTGCCC 1980  
 CTTCTGAGGA TGCTGAGAAC GGTGCTTTTC TTTATAAATG CAAATGGCTA CCGTTTTCAC 2040  
 15 ATAAATTTT GCATGTGCAA AAAAAAAAAA ANAAAAAAAA AAAATCCCGG GGGGGGGCCG 2100  
 GTAACCAATT TGNCCCC 2117

20

(2) INFORMATION FOR SEQ ID NO: 159:

25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2395 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 159:

TGTTCCTTAA TCCTTTTCT AAAAAGGGG GAAATCCCG ATGGATTTTA GGGATTGGTC 60  
 TGGTGTACGC TGTGTTTAT TGCACACCTA AATCCTGATT ATAGGCTTTT CATTTCTCCG 120  
 35 CAAAGCCTTT ATTTTGGCAG TTAAGCCAAA TGTGTTTCC AGAAAGTTAG TTATTTTCTC 180  
 CTCTTTCTTT CCTTTCTTTC CTCCTTTTTT CCCGTCGAC CCCAAACGTT ATTGTCCAAA 240  
 40 CATGACTGGA CAGCAGCTTT TGTTCTTGA CCCTGTAATA TGACAGTCTG CTAATATTGA 300  
 CAGAAGGTGC AGTTTPTGGG TTATAGTCGT GATTTTCGCT AATCAATCAT ATTAGCAGGA 360  
 AAAAAAAGA CTGTTTCTG TTGTAATTGA GTCTTAAGAA AAAGTGGCCC ATAGTTTAGT 420  
 45 GGACAATTTT CAAAGGCTTT AGTACCACCT GTATTTCAAA ATGGGGGACC CAAACTCCCG 480  
 GAAGAAACAA GCTCTGAACA GACTACGTGC TCAGCTTAGA AAGAAAAAG AATCTCTAGC 540  
 50 TGACCAGTTT GACTTCAAGA TGTATATTGC CTTTGTATT CAGGAGAAGA AGAAAAAGTC 600  
 AGCACTTTTT GAAGTGTCTG AGGTTATACC AGTCATGACA AATAATTATG AAGAAAATAT 660  
 CCTGAAAGGT GTGCGAGATT CCAGCTATTC CTTGGAAAGT TCCCTAGAGC TTTTACAGAA 720  
 55 GGATGTGGTA CAGCTCCATG CTCCTCGATA TCAGTCTATG AGAAGGATG TAATTGGCTG 780  
 TACTCAGGAG ATGGATTTC A TTCTTGGCC TCGGAATGAT ATTGAAAAA TCGTCTGTCT 840  
 60 CCTGTTTCT AGGTGGAAAG AATCTGATGA GCCPTTAGG CCTGTTTCAGG CAAATTTGAG 900

	TTTCATCATG GTGACTATGA AAAACAGTTT CTGCATGTAC TGAGCCGCAA GGACAAGACT	960
5	GGAATCGTTG TCAACAATCC TAACCACTCA GTGTTTCTCT TCATTGACAG ACAGCACTTG	1020
	CAGACTCCAA AAAACAAAGC TACAATCTTC AAGTTATGCA GCATCTGCCT CTACCTGCCA	1080
	CAGGAACAGC TCACCCACTG GGGCAGTTGG CACCATAGAG GRTCACCTCC GTCCTTATAT	1140
10	GCCAGAGTAG AGTACTGACC AGCAAAATGG AGAAGATCAG AGAATGCAGC AGCAGTTTTT	1200
	TTTCTTGTTT TCTTACCACT TTATTCTTTC AGAGTTTAAA GAAAATGGAC TCATGCACAG	1260
15	AACACTATGC ATTTTGAAAC TTGTTTCATCC TGGATTTTTT TAAATCATTT TTATCTCAGA	1320
	ACTTAAACAA AAATTAGATG TCGTGCACGG ACTGTGTGAA AGAAGATGCT TTGCATATTT	1380
	GCTGCACTCG ATCAGTATCT TACTAAAAAT GTGAAATGAA AGGACTATTG TACACTGAAA	1440
20	TGCTTAAATG TATCTGAAAG CACAAGGTGA TACTCATTTT TATGGTCTTC CCATTTGTGC	1500
	TGGTTTTTGC CTCTTTGACA TCTGTCATCA GTATTTAGAG GGTGAGAAGT GAATGTAACA	1560
25	GGTATAAATA ACATTTTTTAA AAACAATAAC TTTGCTATAA TCACAGTTGT TCCAGAGCAC	1620
	TGTCAGATAC ATTCTAATGA CCAGAACTGG TTTAAAAAAA GAAAATACAA CCATGGGAAA	1680
	GAAATCTTAA ATGAAAAACG CATCTCATTG TAGGCATTTT TGCCTCATAT TTTACTGGGC	1740
30	CATGTTTGTT TCCTGGTACT CATGTATTTT TTTTTCAG ATCTCTTTCC CCAAGTTGCT	1800
	ATTGTAAGAG TATTCTGCTG CGTGTGGATG CAGTTATACA CATTAAAGCA GATCTGGAGT	1860
35	CTGAAGTAGC TATAAAGCAG CTATAAAACA GAAATACATG CATAGCTGCA GAAACCATGA	1920
	TAGGTAGAGG ACTTTTCITT TGGTTTGTGTT TTGTTTGTGTT TTGTTTGTGTT TTTGGTTTTA	1980
	CAGAGAAGAG ATTTTTATTA CAAAGAAAAA AATTCCAGTG AATTGTGCAG AAATGCTGGT	2040
40	TTTTACACCA TCCTAAAGAA AACTTTTACA AGGGTGTGTT GGAGTAGAAA AAAGGTTATA	2100
	AAGTTGGAAT CTTAAATGTT AAAATTAACC ATTGAGTGTG AAAGTTCTAA AAGCAGAACT	2160
45	CATTTGTGTC AATGAACATA AGGAAAGACT ACTGTATAGG TTTTTTTTTT TTCTCCTTTT	2220
	AAATGAAGAA AAGCTTTGCT TAAGGGTTGC ATACTTTTAT TGGAGTAAAT CTGAATGATC	2280
	CTACTCCTTT GGAGTAAAC TAGTGCTTAC CAGTTTCCAA TTGTATTTAG CTTCTGGTTG	2340
50	GAATTTGAAA AAAAAAGAAA AAAAGAAAAA GAAACCTAA ATAAAATAGG TGAAA	2395

55 (2) INFORMATION FOR SEQ ID NO: 160:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2120 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

60

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 160:

5	CCCCGGATAC CGCCTGACGT AGTGCCAATC ACACCTCTCG CGTCTCGGCG CCTCGGAGGC	60
	TAATGAGGAC GCCTGGCGAA ACGCAGTAAC GGATTTCGG GTGGACCTTC GCTTTACGGC	120
10	TCGTGAGTTC TTCCGCCCAA CCCAGAGGAA GCGGGAGAGC AGTTTACGAC AGCGCCGGTC	180
	GTGTTTACGG CGGCGCCCGC TGCGCGCGCA TGTTCCTCT TTTCTGGTT TCTCAAGAGT	240
	GCTGCTGCTA ACGCGGTCCC CGGCACGCAC CATCTGTTC CATCCCGGCC GGCCGAGGCA	300
15	TTGCAGATTT TGAAGATGG CAAAGTTCAT GACACCCGTG ATCCAGGACA ACCCCTCAGG	360
	CTGGGGTCCC TGTGCGGTTT CCGAGCAGTT TCGGGATATG CCCTACCAGC CGTTCAGCAA	420
20	AGGAGATCGG CTAGGAAAGG TTGCAGACTG GACAGGAGCC ACATACCAAG ATAAGAGGTA	480
	CACAAATAAG TACTCCTCTC AGTTTGGTGG TGAAGTCAA TATGCTTATT TCCATGAGGA	540
	GGATGAAAGT AGCTTCCAGC TGGTGGATAC AGCGCGCACA CAGAAGACGG CCTACCAGCG	600
25	GAATCGAATG AGATTGCCCC AGAGGAACCT CCGCAGAGAC AAAGATCGTC GGAACATGTT	660
	GCAGTTCAAC CTGCAGATCC TGCTTAAGAG TGCCAAACAG AAAGAGAGAG AACGCATTGG	720
30	ACTGCAGAAA AAGTCCAGA AACAATTGG GGTTAGGCAG AAATGGGATC AGAAATCACA	780
	GAAACCCCGA GACTCTTCAG TTGAAGTTGG TAGTGATTGG GAAGTGAAAG AGGAAATGGA	840
	TTTTCTCAG TTGATGAAGA TGCGTACTT GGAAGTATCA GAGCCACAGG ACATTGAGTG	900
35	TTGTGGGGCC CTAGAATACT ACGACAAAGC CTTTGACCGC ATCACCACGA GGAGTGAGAA	960
	GCCACTGCGG ASATNCAAGC GCATCTTCCA CACTGTACAC ACCACAGACG ACCCTGTCTAT	1020
40	CCGCAAGCTG GCAAAAATC AGGGGAATGT GTTTGCCACT GATGCCATCC TGGCCACGCT	1080
	GATGAGCTGT ACCCGCTCAG TGTATTCTTG GGATATTGTC GTCCAGAGAG TTGGGTCCAA	1140
	ACTCTTCTTT GACAAGAGAG ACAACTCTGA CTTTGACCTC CTGACAGTGA GTGAGACTGC	1200
45	CAATGAGCCC CCTCAAGATG AAGGTAATTC CTTCAATTCA CCCCACAACC TGGCCATGGA	1260
	GGCAACCTAC ATCAACCACA ATTTCTCCA GCAGTGCTTG AGAATGGGA AGGAAAGATA	1320
50	CAACTTCCCC AACCCAAACC CGTTTGTTGA GGACGACATG GATAAGAATG AAATCGCCTC	1380
	TGTTGCGTAC CGTTACCGCA GTGGNAAGCT TGGAGATGAT ATTGACCTTA TTGTCCGTTG	1440
	TGAGCACGAT GCGGTCATGA CTGGAGCCAA CGGGGAAGTG TCCTTCATCA ACATCAAGAC	1500
55	ACTCAATGAG TGGGATTCCA GGCAGTGTAA TGGCGTTGAC TGGCGTCAGA AGCTGGACTC	1560
	TCAGCGAGGG GCTGTCTATT CCACGGAGCT GAAGAACAAC AGCTACAAGT TGGCCCGGTG	1620
60	GACCTGCTGT GCTTTGCTGG CTGGATCTGA GTACCTCAAG CTGGTTATG TGTCTCGGTA	1680



CCACGTGAAA GACTCCTCAC GCCACGTCAT CCTAGGCACC CAGCAGTTCA AGCCTAATGA 1740  
 GTTTGCCAGC CAGATCAACC TGAGCGTGA GAATGCCTGG GGCATTTTAC GCTGCGTCAT 1800  
 5 TGACATCTGC ATGAAGCTGG AGGAGGGCAA ATACCTCATC CTCAAGGACC CCAACAAGCA 1860  
 GGTCATCCGT GTCTACAGCC TCCCTGATGG CACCTTCAGC TCTGATGAAG ATGAGGAGGA 1920  
 AGAGGAGGAG GAAGAAGAGG AAGAAGAAGA GGAAGAACT TAAACCACTG ATGTGGAGCT 1980  
 10 GGAGTTTGTG CTTCCACCGA GACTACGAGG GCCTTTGATG CTTAGTGGAA TGTGTGTCTA 2040  
 ACTTGCTCTC TGACATTTAG CAGATGAAAT AAAATATATA TCTGTTTAGT CTTAAAAAAA 2100  
 15 AAAAAAAAAA AAAAAAAAAAN 2120

20 (2) INFORMATION FOR SEQ ID NO: 161:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 900 base pairs  
 (B) TYPE: nucleic acid  
 25 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 161:

30 GGAAGCTGAA GTCCTTCCAG ACCAGGGACA ACCAGGGCAT TCTCTATGAA GCTGCACCCA 60  
 CCTCCACCCT CACCTGTAC TCAGGACCAC AGAAGCAAAA GTTCTCACTC AAACCTGGATG 120  
 CCAAGGATGG GCGCTGTTC AATGAGCAGA ACTTCTTCCA GCGGGCCGCC AAGCCTCTGC 180  
 35 AAGTCAACAA GTGGAAGAAG CTGTACTCGA CCCCACTGCT GGCCATCCCT ACCTGCATGG 240  
 GTTTCGGTGT TCACCAGGAC AAATACAGGT TCTTGGTGT ACCCAGCCTG GGGAGGAGCC 300  
 40 TTCAGTCGGC CCTGGATGTC AGCCCAAAGC ATGTGCTGTG CAGAGAGGTC TGTGCTGCAG 360  
 GTGGCCTGCC GGCTGCTGGA TGCCCTGGAG TTCCTCCATG AGAATGAGTA TGTTCATGGA 420  
 AATGTGACAG CTGAAAATAT CTTGTGGAT CCAGAGGACC AGAGTCAGGT GACTTTGGCA 480  
 45 GGCTATGGCT TCGCMTCCG CTATTGCCA AGTGGCAAAC ACGTGGCCTA CGTGAAGGC 540  
 AGCAGGAGCC CTCACGAGG GGACCTTGAG TTCATTAGCA TGGACCTGCA CAAGGGATGC 600  
 50 GGGCCCTCCC GCGCRGCGA CCTCCAGAGC CTGGGCTACT GCATGCTGAA GTGGCTCTAC 660  
 GGGTTTCTGC CATGGACAAA TTGCCCTCCC AAMAMTGAG ACATCATGAA GCAAAAACAG 720  
 AAGTTTGTG ATAAGCCGGG GCCCTTCGTG GGACCTGCG GTCACCTGAT CAGGCCCTCA 780  
 55 GAGACCTCG AGAAGTACCT GAAGGTGGTG ATGGCCCTCA CGTATGAGGA GAAGCCGCCC 840  
 TACGCCATGC TGAGGAACAA CCTAGAAGCT TTGCTGCAGG ATCTGCGTGT GTCTCCATAT 900

60

## (2) INFORMATION FOR SEQ ID NO: 162:

5

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1003 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 162:

GGCACGAGAT GAGGGGCACC CAGTGCTTCT AGGGCAGGCT GGGTGGTGGT CCCCTAGGTA 60  
15 TCAGCCTCTC TTACTGTACT CTCGGGAAT GTTAACCTTT CTATTTTCAG CCTGTGCCAC 120  
CTGTCTAGGC AAGCTGGCTT CCCCATTTGC CCCTGTGGGT CCACAGCAGC GTGGCTGCCC 180  
20 CCCAGGGCCA CCGCTTCTTT CTTGATCCTC TTTCCTTAAC AGTGACTTGG GCTTGAGTCT 240  
GGCAAGGAAC CTTGCTTTTA GCTTCACCAC CAAGGAGAGA GGTGACATG ACCTCCCCGC 300  
CCCCTACCA AGGCTGGGAA CAGAGGGGAT GTGGTGAGAG CCAGGTTCTT CTGGCCCTCT 360  
25 CCAGGTGTT TTCCACTAGT CACTACTGTC TTCTCCTTGT AGCTAATCAA TCAATATTCT 420  
TCCCTTGCCT GTGGGCAGTG GAGAGGCTGC TGGGTGTACG CTGCACCTGC CCACTGAGTT 480  
GGGAAAGAG GATAATCAGT GAGCACTGTT CTGCTCAGAG CTCCTGATCT ACCCCACCCC 540  
30 CTAGGATCCA GGACTGGGTC AAAGCTGCAT GAAACCAGGC CCTGGCAGCA AACCTGGGAA 600  
TGGCTGGAGG TGGGAGAGAA CCTGAATTC TCTTTCCTC TCCTCCTCC AACATTACTG 660  
35 GAACTCTATC CTGTTAGGAT CTTCTGAGCT TGTTCCTTG CTGGGTGGGA CAGAGGACAA 720  
AGGAGAAGGG AGGGTCTAGA AGAGGCAGCC CTTCTTTGTC CTCTGGGTA AATGAGCTTG 780  
ACCTAGAGTA AATGGAGAGA CCAAAGCCT CTGATTTTAA ATTTCATAA AATGTTAGAA 840  
40 GTATATATAT ACATATATAT ATTTCTTTAA ATTTTGTAGT CTTTGATATG TCTAAAAATC 900  
CATTCCTCT GCCCTGAAGC CTGAGTGAGA CACATGAAGA AAACGTGTT TCATTAAAG 960  
45 ATGTTAATTA AATGATTGAA ACTTGAAAAA AAAAAAAAAA AAA 1003

50

## (2) INFORMATION FOR SEQ ID NO: 163:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2196 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

55

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 163:

60

AAGAAGCGGC ACACGGATGT GCAGTCTAC ACAGAAGTGG GAGAGATAAC CACGGACTTG 60

	GGGAAACATC AGCATATGCA TGACCGAGAT GACCTCTATG CTGAGCAGAT GGAACGAGAA	120
5	ATGAGGCACA AACTGAAAAC AGCCTTTAAA AATTTCATTG AGAAAGTAGA GGCTCTAACT	180
	AAGGAGGAAC TGGAATTTGA AGTGCCCTTTT AGGGACTTGG GATTTAACGG AGCTCCCTAT	240
	AGGAGTACCT GCCTCCTTCA GCCCACTAGT AGTGCGCTGG TAAATGCTAC GGAATGGCCA	300
10	CCTTTGTGG TGACATTGGA TGAGGTAGAG CTGATCCACT TTRAGCGGGT CCAGTTTCAC	360
	CTGAAGAACT TTGATATGGT AATCGTCTAC AAGGACTACA GCAAGAAAGT GACCATGATC	420
15	AACGCCATTG CTGTAGCCTC TCTTGACCCC ATCAAGGAAT GGTGGAATTC CTGCGACCTG	480
	AAATACACAG AAGGAGTACA GTCCCTCAAC TGGACTAAAA TCATGAAGAC CATTTGTTGAT	540
	GACCCGTAGG GCTTCTTCGA ACAAGGTGGC TGGTCTTTCC TGGAGCCTGA GGGTGAGGGG	600
20	AGTGATGCTG AAGAAGGGGA TTCAGAGTCT GAAATTGAAG ATGAGACTTT TAATCCTTCA	660
	GAAGATGACT ATGAAGAGGA AGAGGAGGAC AGTGATGAAG ATTATTTCATC AGAAGCAGAA	720
25	GAGTCAGACT ATTCTAAGGA GTCATTGGGT AGTGAAGAAG AGAGTGGAAG GGATTGGGAT	780
	GAAGTGAGG AAGAAGCCCG AAAAGCGGAC CGAGAAAGTC GTTACGAGGA AGAAGAAGAA	840
	CAAAGTCGAA GTATGAGCCG GAAGAGGAAG GCATCTGTGC ACAGTTCGGG CCGTGGCTCT	900
30	AACCGTGGTT CCAGACACAG CTCTGCACCC CCCAAGAAAA AGAGGAAGTA ACTTCTGAAC	960
	TTTGCCCCG AGCTCCATTG TTCTCCAGC CAACCCCTGA AAATTTTACA TGACATAGAA	1020
35	ACTGTATTTT TCCTTTGGTT TTCAATTGAA GTTTTGCCAT TTGTGTTTAT GGGTTTAGGG	1080
	GGCCATTGT GTGGACCAAT CTACTCGGGG AATTCCAGGC CCACCAGGAC ACGTGCCAAT	1140
	GGCCCCATTG AGATGGCAAG GGAGGAGGTG TTCTTGAAGA CAGGAGGAGG CTCCCGCTGT	1200
40	TAATAAATAT TGTTTCATTG TTCTCTCTTC CTGTCACTT CTGCCAAGAC ATTGATGGCT	1260
	TCTGACATCT TATTTGGTGT CTCAAAGCTG TATTTCCAAG ACAGTGGTAC AAGGTGACCC	1320
45	TTAATTACCC GTATCATGGT TCTTGACCAG CACATTCAAT CCTCCAACCT ACCCTACTGC	1380
	CATGACCTTC CGCACATCTC TAAGTTTAT CTTTGCAATA CTCAGGTTC TCGGAAATTT	1440
	GCTAATGGTT GTGATAAACC ATACAGCTTG AGCCAGTGAG GCAGATTGGG CTGGTGCCCT	1500
50	CGTCTGAGTT TTCCTGCTTT CCTGCCCTGT GCAGATTCTG AGGTATATCT GCTGCCTTGG	1560
	AAGACATAAG AAGCAGTGAT ACTCCCTGGC TCGGTATTT TCTCCATACA ATGCACACAT	1620
55	GGTACAATGA TAGAAGGCAA AATTGCCACT GTCTTCTTTT TTTTCTCATA TATCTAAGGA	1680
	AGATATATCA GGTGTGCCT CATGTACCGC TTCTAGTGAA ATGTAGAGGA AGGCTCAAAG	1740
	GAGTCAACAT TTAGATCTGG AAGGGACAAG TCATGCCTTG GGCCTAGAAT ACCCTGATGA	1800
60	GAAAAGAGAA GAGGAAGGGA GGCCATATCT ACAACANCAN CCTCTCGGCA CTGCTGCTCC	1860

TTATTTTAAC TTGTCTTGC ATTGTCTGT ATTTATCACA GTTCTGTGTG AACAGCTTTT 1920  
5 CAAGTATTTG GGGAGTTTAT CTTGCCATCC TCCCCTTCTG GTTCTCTGCA CCCACCTGTC 1980  
CCACTGCAGT TCCTTCCGTG CTCTGTGACT TTAAGAGAAG AAGGGGGGAG GGGTCCCGGA 2040  
TTTTATGTTT GTTTGTTTTT TCTCCTTAGC AGTAGGACTT GATATTTTCA ATTTTGAAG 2100  
10 AACTAAAAGA TGAATAAACT GGGTTTTTTT TGTGTGTTGT TTTGTAAAA AAAAAAAAAA 2160  
AAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAA 2196

15

(2) INFORMATION FOR SEQ ID NO: 164:

(i) SEQUENCE CHARACTERISTICS:  
20 (A) LENGTH: 1945 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 164:

GCACAGAGTC GGGCGGACGG ACAGGGAGAG GAGGAGAGGG GGTCTGCGCG CGGCCGCTAC 60  
CCAGAAGCCA GCGGACGGCA GCACGGAGTG GGCTGTCCCC GAGCCCAGCC CCGAGCGAGC 120  
30 CCCCCCCCCG CCCCCGMAGG ACGCGCCTYC CAGCCAGCCC GACTYCTAGG AGGAGGGGAG 180  
GCGGAAAGC AGCTCAAGCC TCACCCACCG CCCTGCCCCC AGCCCCGCCA CTCCCAGGCT 240  
35 CCTCGGGACT CGCGGGTCC TCCTGGGAGT CTCGGAGGGG ACCGGCTGTG CAGACGCCAT 300  
GGAGTGGTG CTGGTCTTCC TCTGCAGCCT GCTGGCCCCC ATGGTCTCTG CCAGTGCAGC 360  
TGAAAAGGAG AAGGAAATGG ACCCTTTTCA TTATGATTAC CAGACCCCTGA GGATGGGGG 420  
40 ACTGGTGTTC GCTGTGGTCC TCTCTCGGT TGGGATCCTC CTTATCCTAA GTCGCAGGTG 480  
CAAGTGCAGT TTCAATCAGA AGCCCCGGGC CCCAGGAGAT GAGGAAGCCC AGGTGGAGAA 540  
45 CCTCATCACC GCCAATGCAA CAGAGCCCCA GAAAGCAGAG AACTGAAGTG CAGCCATCAG 600  
GTGGAAGCCT CTGGAACCTG AGCGGGCTGC TTGAACCTTT GGATGCAAAT GTCGATGCTT 660  
AAGAAAACCG GCCACTTCAG CAACAGCCCT TTCCCCAGGA GAAGCCAAGA ACTTGTGTGT 720  
50 CCCCCACCCT ATCCCCCTA ACACCATTC TCCACCTGAT GATGCAACTA ACACTTGCCT 780  
CCCCACTGCA GCCTGCGGTC CTGCCACCT CCCGTGATGT GTGTGTGTGT GTGTGTGTGT 840  
55 GTGACTGTGT GTGTTTGCTA ACTGTGGTCT TTGTGGCTAC TTGTTTGTGG ATGGTATTGT 900  
GTTTGTAGT GAAGTGTGGA CTCGCTTTC CAGGCAGGGG CTGAGCCACA TGGCCATCTG 960  
CTCCTCCCTG CCCCCGTGGC CCTCCATCAC CTCTGCTCC TAGGAGGCTG CTTGTGCCCC 1020  
60

	GAGACCAGCC CCCTCCCTG ATTTAGGGAT GCGTAGGGTA AGAGCACGGG CAGTGGTCTT	1080
	CAGTCGTCTT GGGACCTGGG AAGGTTTGCA GCACTTTGTC ATCATTCTTC ATGGACTCCT	1140
5	TTCACTCCTT TAACAAAAAC CTTCCTTCCT TATCCCACCT GATCCCAGTC TGAAGGTCTC	1200
	TTAGCAACTG GAGATACAAA GCAAGGAGCT GGTGAGCCCA GCGTTGACGT CAGGCAGGCT	1260
10	ATGCCCTTCC GTGGTTAATT TCTTCCCAGG GGCTTCCACG AGGAGTCCCC ATCTGCCCCG	1320
	CCCCCTCACA GAGCGCCCGG GGATTCCAGG CCCAGGGCTT CTACTCTGCC CCTGGGGAAT	1380
	GTGTCCCTG CATATCTTCT CAGCAATAAC TCCATGGGCT CTGGGACCTT ACCCTTCCA	1440
15	ACCTTCCCTG CTCTGAGAC TTCAATCTAC AGCCAGCTC ATCCAGATGC AGACTACAGT	1500
	CCCTGCAATT GGTCTCTGG CAGGCAATAG TTGAAGGACT CCTGTTCCGT TGGGGCCAGC	1560
20	ACACCGGGAT GGATGGAGGG AGAGCAGAGG CCTTTGCTT TCTGCCTACG TCCCCTTAGA	1620
	TGGGCAGCAG AGGCAACTCC CGCATCCTTT GCTCTGCCTG TCRGTGGTCA GAGCGGTGAG	1680
	CGAGGTGGGT TGGAGACTCA GCAGGCTCCG TGCAGCCCTT GGAACAGTG AGAGGTTGAA	1740
25	GGTCATAACG AGAGTGGGAA CTCAACCCAG ATCCCGCCCC TCCTGTCTC TGTGTPCCCG	1800
	CGGAAACCAA CCAACCGTG CGCTGTGACC CATTGCTGTT CTCTGTATCG TGATCTATCC	1860
30	TCAACAACAA CAGAAAAAAG GAATAAATA TCCTTTGTTT CCTAGTGAAA AAAAAAAAAA	1920
	AAAAAAAAA AAAAAAAAAA CTCGA	1945

35

(2) INFORMATION FOR SEQ ID NO: 165:

## (i) SEQUENCE CHARACTERISTICS:

40

- (A) LENGTH: 2933 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 165:

45

	GGGTCGACCC ACGCGTCGG CAGCGTCGT TTGAGTGGT GTCGCCGTG CCCCCTCCCG	60
	GATCAGGAGC CAGTGTATAC CGCCGCCCA CCGCTTGGT GCCGCTAGAG GAAACGAGAA	120
50	GGAGGCCGCC TCGGTTTGT CGCCGAGCT CGCCMCYGY CYGGRAGAGC CGAGCCCCG	180
	CCAGTCGGT CGYTGCCAC CSCTCGTAGC CGTTACCCG GGGCGCCAC AGCCGCGGC	240
55	CGGAGAGGC GCGGCCCATG GCTCTGGAG CCGATTCAA AGGTGATGAC CTATCAACAG	300
	CCATTCTCAA ACAGAAGAAC CGTCCAATC GGTAAATGT TGATGAAGCC ATCAATGAGG	360
	ACAACAGTGT GGTGCTCTG TCCCAGCCA AGATGGATGA ATTGCAGTGT TTCCGAGGTG	420
60	ACACAGTGT GCTGAAAGGA AAGAAGAGC GAGAAGCTGT TTGCATCGTC CTTTCTGATG	480

	ATACTTGTTT	TGATGAGAAG	ATTCGGATGA	ATAGAGTTGT	TCGGAATAAC	CTTCGTGTAC	540
5	GCCTAGGGGA	TGTCATCAGC	ATCCAGCCAT	GCCCTGATGT	GAAGTACGGC	AAACGTATCC	600
	ATGTGCTGCC	CATTGATGAC	ACAGTGAAG	GCAITTAAGG	TAATCTCTTC	GAGGTATACC	660
	TTAAGCCGTA	CTTCCTGGAA	GCGTATCGAC	CCATCCGGAA	AGGAGACATT	TTTCTTGTC	720
10	GTGGTGGGAT	GCGTCTGTG	GAGTTCAAAG	TGGTGGAAAC	AGATCCTAGC	CCTTATTGCA	780
	TTGTGCTCC	AGACACAGTG	ATCCACTGCG	AAGGGAGGCC	TATCAAACGA	GAGGATGAGG	840
15	AAGAGTCCTT	GAATGAAGTA	GGGTATGATG	ACATTTGGTG	CTGCAGGAAG	CAGCTAGCTC	900
	AGATAAAGGA	GATGGTGGAA	CTGCCCTGA	GACATCCTGC	CCTCTTTAAG	GCAATTGGTG	960
	TGAAGCCTCC	TAGAGGAATC	CTGCTTTACG	GACCTCCTGG	AACAGGAAAG	ACCTTGATTG	1020
20	CTCGAGCTGT	AGCAAATGAG	ACTGGAGCCT	TCTTCTTCTT	GATCAATGGT	CCTGAGATCA	1080
	TGAGCAAATT	GGCTGGTGAG	TCTGAGAGCA	ACCTTCGTAA	AGCCTTTGAG	GAGGCTGAGA	1140
25	AGAATGCTCC	TGCCATCATC	TTCATTGATG	AGCTAGATGC	CATCGCTCCC	AAAAGAGAGA	1200
	AAACTCATGG	CGAGGTGGAG	CGGCGCATTG	TATCACAGTT	GTTGACCTTC	ATGGATGGCC	1260
	TAAAGCAGAG	GGCACATGTG	ATTGTTATGG	CAGCAACCAA	CAGACCCAAC	AGCATTGACC	1320
30	CAGCTCTACG	GCGATTTGGT	CGCTTTGACA	GGGAGGTAGA	TATTGGAAIT	CCTGATGCTA	1380
	CAGGACGCTT	AGAGATTCTT	CAGATCCATA	CCAAGAACAT	GAAGCTGGCA	GATGATGTGG	1440
35	ACCTGGAACA	GTAGCCAATG	AGACTCAOGG	GCATGTGGGT	GCTGACTTAG	CAGCCCTGTG	1500
	CTCAGAGGCT	GCTCTGCAAG	CCATCCGCAA	GAAGATGGAT	CTCATTGACC	TAGAGGATGA	1560
	GACCAITGAT	GCCGAGGTCA	TGAACCTCTT	AGCAGTTACT	ATGGATGACT	TCCGGTGGGC	1620
40	CTTGAGCCAG	AGTAACCCAT	CAGCACTGCG	GGAAACCGTG	GTAGAGGTGC	CACAGGTAAC	1680
	CTGGGAAGAC	ATCGGGGGCC	TAGAGGATGT	CAAACGTGAG	CTACAGGAGC	TGGTCCAGTA	1740
45	TCTGTGGAG	CACCCAGACA	AATTCTTGAA	GTITGGCATG	ACACCTTCCA	AGGGAGTTCT	1800
	GTTCATGGA	CCTCTGGCT	GTGGGAAAAC	TTTGTGGGCC	AAAGCCATTG	CTAATGAATG	1860
	CCAGGCCAAC	TTCATCTCCA	TCAAGGTCC	TGAGCTGCTC	ACCATGTGGT	TTGGGGAGTC	1920
50	TGAGGCCAAT	GTCAGAGAAA	TCTTTGACAA	GGCCCGCCAA	GCTGCCCCCT	GTGTGCTATT	1980
	CTTTGATGAG	CTGGATTGGA	TTGCCAAGGC	TCGTGGAGGT	AACATTGGAG	ATGGTGGTGG	2040
55	GGCTGCTGAC	CGAGTCATCA	ACCAGATCCT	GACAGAAATG	GATGGCATGT	CCACAAAAAA	2100
	AAATGTGTTT	ATCATTTGGG	CTACCAACCG	GCCTGACATC	ATTGATCCTG	CCATCCTCAG	2160
	ACCTGGCCGT	CTTGATCAGC	TCATCTACAT	CCACTTCCT	GATGAGAAGT	CCCGTGTGTC	2220
60	CATCCTCAAG	GCTAACCTGC	GCAAGTCCCC	AGTTGCCAAG	GATGTGGACT	TGGAGTTCTT	2280

GGCTAAATG ACTAATGGCT TCTCTGGAGC TGACCTGACA GAGATTGGCC AGCGTGCTTG 2340  
 CAAGCTGGCC ATCCGTGAAT CCATCGAGAG TGAGATTAGG CGAGAACGAG AGAGGCAGAC 2400  
 5 AAACCCATCA GCCATGGAGG TAGAAGAGGA TGATCCAGTG CCTGAGATCC GTCGAGATCA 2460  
 CTTTGAAGAA GCCATGGGCT TTGCGCGCCG TTCTGTCACT GACAAATGACA TTCGGAAGTA 2520  
 10 TGAGATGTTT GCCCAGACCC TTCAGCAGAG TCGGGGCTTT GGCAGCTTCA GATTCCTTTC 2580  
 AGGGAACCAG GGTGGAGCTG GCCCCAGTCA GGGCAGTGA GCGGCACAG GTGCGAGTGT 2640  
 ATACACAGAA GACAATGATG ATGACCTGTA TGGCTAAGTG GTGGTGGCCA GCGTGCAGTG 2700  
 15 AGCTGGCCTG CCTGGACCTT GTTCCTTGGG GGTGGGGGCG CTGCCCAGG AGAGGGACCA 2760  
 GGGGTGCGCC CACAGCCTGC TCATTCTCC AGTCTGAACA GTTCAGCTAC AGTCTGACTC 2820  
 20 TGGACAGGGG GTTCTGTGTG CAAAAATACA AAACAAAAGC GATAAAATAA AAGCGATTTT 2880  
 CATTTGGTAA AAAAAAAAAA AAAAAAAT CCGGGGGGGG GCCCGAACCA TTT 2933

25

(2) INFORMATION FOR SEQ ID NO: 166:

30 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2243 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 166:

TCGGAGAGCC GCGGGGCGNG CGCTCTCGG CCAGGAAGCG CCTCTGGAC GCGTGTNACC 60  
 GATGCCCAGA AGTGGCCTTG GGCTGGGGAT CACCATAGCT TTTCTAGCTA CGCTGATCAC 120  
 40 GCAGTTTCTC GTGTATAATG GTGTCTATCA GTATACATCC CCAGATTTCC TCTATATTCG 180  
 TTCTTGCTC CCTGTATAT TTTTCTCAGG AGGCGTCAG GTGGGAACA TAGGACGACA 240  
 45 GTTAGCTATG GGTGTTCTG AAAAGCCCCA TAGTGATTGA GTCTTCAAAA CCACCGATTC 300  
 TGAGAGCAAG GAAGATTTTG GAAGAAAATC TGA CTGTGGA TTATGACAAA GATTATCTTT 360  
 TTTCTTAAGT AATCTATTTA GATCGGGCTG ACTGTACAAA TGA CTCTGG AAAAACTCT 420  
 50 TCACCTAGTC TAGAATAGGG AGGTGGAGAA TGATGACTTA CCCTGAAGTC TTCCCTTGAC 480  
 TGCCCGCACT GCGCCTGTC TGTGCCCTGG AGCATTCTGC CCAGGCTACG TGGGTTGAG 540  
 55 CAGGTGGCAG CTTCCCAAGT ATTCGATTTC ATTCATGTGA TTAACAAG TTGCCATATT 600  
 TCAAAGCCTT GAAC TAAGAC TCAATTACCA ACCCGCAGTT TTGTGTCACT GCCCAAAGGA 660  
 60 GGTAGGTGA TGGTGCTTAA CAAACATGAA GTATGGTGTA ATAGGAATAA TATTTATCCA 720

	AAAGATTTTT AAAAATAGGG CTGTGTTTAA AAAAAAAAAAC AAAACARGAA AAGCAGCAGT	780
	GATTATAGAG AGGTCACACT CTAAGTGGGG TCGCGGCGTG GCCACGCTTC ACGGTCACGC	840
5	TCGTCCGTCC TGCAGTGGCG TGTTTACATG GTCACACGTG TGTGTATCAC CAGTGGGTCA	900
	ACTGCTTGTC ATTCTCCCG TGGCAGTTTG TGTAGACAAT CTTACTGAGC AAAAGGCAAT	960
10	GAAAAGTCTT GGTCCACACA CTGCGATATA TTGGAATTTT CACCTCAGTT TATGAAGTTT	1020
	ATTTCGAAAT CCATAGTCAT CTAAGAATGA ATACCTGTCT GCCATGTATT TCAATCTTAG	1080
	TGAGCCAAAA TTGTTTGTGTT GTTACTACAG AATAGAGATG ACTGTTTTTT GCCACAGCCC	1140
15	TATGGRATTT GCAATCTGTG ATTGCCTTGT AAAAAGGAGA GTGCATATGG CACTGCATTA	1200
	AACGTGTGGT GTTCTAGTC AATGATATTG GTGAGCACAA TGTATTCATT TAATGGCATA	1260
20	GACCATACCA GACCTAATTT GCAAGTATTG GGTCTTAAAC TTCAAGTGCA ATGTATATGA	1320
	AAACCAATCT GAGCCTTGTA TCTCTTAAAT ATTTATTTTT TTTAACGTGT GAGATGTTTG	1380
	AGAGAAGGTT CTCCATTTCAT TTCAGTGTG CCTGGAGGAA ACTCGGCAAT GATTTCTTTC	1440
25	AGTTGTGAAG TTCTTTTCGT GTTACACCCT CCACTGAACC CTCAACCTTC GAAATACTCC	1500
	AGTTTGTGG GTTTGGTCAT TTTTACTTAT AAATTTACCT TTTTGTATTT TGCAATTTAC	1560
30	ATGTGTTTGG TTTGTTTTAA ATTCTGTGAA AGTGGCTTGA TTAAAAGACT CCTTTTAAAT	1620
	GGAAGCCACC AGTCAGCAGA ATGGAAGCTT AGAGGAACTT GCCTGTGAGC GCTGGTCTTT	1680
	GTGTTTGGTT TTGTGATGTA ACGATCTTTG CTGGGTTTTT TTGCTTTGTT TTGAGGGAAA	1740
35	TGTCCTGGAG TAAATTTTAA GTTCTGGAG TTAATTTGTT TTACAGGAAT TTTGTTTTTT	1800
	AAAAAATAG GATCATTTCTG AACTTTGGAA TGACCCCTT ATATATTTTC TGAAAATGAA	1860
40	AACAGTTACA TGAAAAAAT TTCCAATGAA GATGTCAGCA TTTTATGAAA AACCAGAAGT	1920
	TATTAGATGA AAGCAGCGAG TGAATCTTTA AACAGACTT GATCACGCAC ACACAATAAG	1980
	TCTTTCTCTC CGAAACCGGA AGTAAATCTA TATCTGTTAG AAATAATGTA GCCAAAAGAA	2040
45	TGTAAATTG AGGATTTTTT TGCCAATAGT TTATAGAAAA TATATGAACC AAAGTGATT	2100
	GAGTTGTAA AAATGTAAAA TAGTATGAAC AAAATTTGCA CTCTACCAGA TTTGAACATC	2160
50	TAGTGAGGTT CACATTCATA CTAAGTTTTC AACATTGIGT TCTTTTTGCA TTCATTTTTT	2220
	ACTTTTATTA AAGGTTCAAA ACC	2243

55

(2) INFORMATION FOR SEQ ID NO: 167:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1816 base pairs

(B) TYPE: nucleic acid

60



(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 167:

5	GGTGGGNAGC TTINAATTTTC CCTTACWGG GGCGCTNTAA GGGGAAACCT TCCCGGAATT	60
	TTCCGGTTCGA CCCACGCGTC CGGCCAGCCT AGGAGAAGAA GTTCGTAGTC CCAGAGGTGA	120
10	GGCAGGAGGC GGCAGTTTCT GCGGGGTGAG GCGGGAGCTG AAGTGACAGC GGAGGCGGAA	180
	GCAACGGTCG GTGGGGCGGA GAAGGGGGCT GCGCCAGGA GGAGGAGGAA ACCCTTCCGA	240
	GAAAACAGCA ACAAGCTGAG CTGCTGTGAC AGAGGGGAAC AAGATGGCGG CGCCGAAGGG	300
15	GAGCCTCTGG GTGAGGACCC AACTGGGGCT CCGCGCGCTG CTGCTGCTGA CCATGGCCTT	360
	GGCCGGAGGT TCGGGGACCG CTTCGGCTGA AGCATTTGAC TCGGTCCTGG GTGATACGGC	420
20	GTCTTGCCAC CGGGCCTGTC AGTTGACCTA CCCCTTGCAC ACCTACCCTA AGGAAGAAGA	480
	GTTGTACGCA TGTACAGAGG GTTGCAGGCT GTTTTCAATT TGTAGTTTG TGGATGATGG	540
	AATTGACTTA AATCGAACTA AATTGGAATG TGAATCTGCA TGTACAGAAG CATATTCCCA	600
25	ATCTGATGAG CAATATGCTT GCCATCTTGG RTGCCAGAAT CAGCTGCCAT TCGCTGAACT	660
	GAGACAAGAA CAACTTATGT CCCTGATGCC AAAAATGCAC CTACTCTTTC CTCTAACTCT	720
30	GGTGAGGTCA TTCTGGAGTG ACATGATGGA CTCCGCACAG AGCTTCATAA CCTCTTCATG	780
	GACTTTTATAT CTTCAAGCCG ATGACGGAAA AATAGTTATA TTCCRGTTCTA AGCCCAGRAA	840
	TCCCAGGTAC GCACCACATT TGGAGCCAGG AGCCCTACCA AATTTGRGRG RAWCMCTCTT	900
35	AAGCAAAATG TCCNTCAKMT CGSMAATGAG AAATTCACAA GCGCACAGGA ATTTTCTTGA	960
	AGATGGAGAA AGTGATGGCT TTTTAAGATG CCTCTCTCTT AACTCTGGGT GGATTTTAAC	1020
40	TACAACCTCTT GTCCCTCTCGG TGATGGTATT GCTTTGGATT TGTGTGCAA CTTGTTGCTA	1080
	CACGCTGTTG GACGCAGTAT AGTTTCCCTC TGAGAAGCTG AGTATCTATG GTGACTTGGA	1140
	GTTTATGAAT GAACAAAAGC TAAACAGATA TCCAGCTTCT TCTCTGTGG TTGTTAGATC	1200
45	TAAACTGAA GATCATGAAG AAGCAGGGCC TCTACCTACA AAAGTGAATC TTGCTCATTC	1260
	TGAAATTTAA GCATTTTCTT TTTAAAAGAC AAGTGTAATA GACATCTAAA ATTCCACTCC	1320
50	TCATAGAGCT TTTAAAATGG TTTTATTGGA TATAGGCCTT AAGAAATCAC TATAAAATGC	1380
	AAATAAAGTT ACTCAAATCT GTGAAAAAAA AAAAAAAAAA AAAAAAAAC TCGAGGGGGG	1440
	GCCCGTTACC AAKTCGCCCT ATWGTGADTB GTATTMTTAT TTTACTAATA TCTGTAGCTA	1500
55	TTTTGTTTTT KGCTTKGGTT ATKGTTTTTY TCCCTTYTCT WAGCTATRAG CTGATCATKG	1560
	CYSCTTCTCA CCTCCTGCCA TGATACTGTC AGTTACCTTA GTTAACAAGC TGAATATTTA	1620
60	GTAGAAATGA TGCTTCTGCT CAGGAATGGC CCACAAATCT GTAATTGAA ATTTAGCAGG	1680

AAATGACCTT TAATGACACT ACATTTTCAG GAACTGAAAT CATTAAAAAT TTATTTGAAT 1740  
AATTATGTGC TGAAAAA AAAA AAAA AMWMRARASK RRWWACTCGA GGGGGGCCCC 1800  
5 GGTACCCNAT TCGCCG 1816

10

(2) INFORMATION FOR SEQ ID NO: 168:

(i) SEQUENCE CHARACTERISTICS:  
15 (A) LENGTH: 945 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 168:

AGAAACCGTT GATGGGACTG AGAAACCAGA GTTAAACCT CTTTGGAGCT TCTGAGGACT 60  
CAGCTGGAAC CAACGGGCAC AGTTGGCAAC ACCATCAACT TCTCCCAAGC AGAGAAACCC 120  
25 GAACCCACCA ACCAGGGGCA GGATAGCCTG AAGAAACATC TACACGCAGA AATCAAAGTT 180  
ATTGGGACTA TCCAGATCTT GTGTGGCATG ATGGTATTGA GCTTGGGGAT CATTTTGGCA 240  
TCTGCTTCCT TCTCTCCAAA TTTTACCCAA GTGACTTCTA CACTGTTGAA CTCTGCTTAC 300  
30 CCATTCATAG GACCCCTTTT TTTTATCATC TCTGGCTCTC TATCAATCGC CACAGAGAAA 360  
AGGTTRACCA AGCTTTTGGT GCATAGCAGC CTGGTTGGAA GCATTCTGAG TGCTCTGTCT 420  
35 GCCCTGGTGG GTTTCATTAT CCTGTCTGTC AACAGGCCA CCTTAAATCC TGCCTCACTG 480  
CAGTGTGAGT TGGACAAAAA TAATATACCA ACAAGAAGTT ATGTTTCTTA CTTTATCAT 540  
GATTCACTTT ATACCAGGA CTGCTATACA GCCAAAGCCA GTCTGGCTGG AWCTCTCTCT 600  
40 CTGATGCTGA TTTGCACTCT GCTGGAATTC TGCCTAGCTG TGCTCACTGC TGTGCTGCGG 660  
TGGAACAGG CTTACTCTGA CTTCCTGGG AGTGTACTTT TCCTGCCTCA CAGTTACATT 720  
45 GGTAATTCTG GCATGTCTC AAAAATGACT CATGACTGTG GATATGAAGA ACTATTGACT 780  
TCTTAAGAAA AAAGGGAGAA ATATTAATCA GAAAGTTGAT TCTTATGATA ATATGGAAAA 840  
GTTAACCATT ATAGAAAAGC AAAGCTTGAG TTTCTTAAAT GTAAGCTTTT AAAGTAATGA 900  
50 ACATTAAAAA AAACCATTAT TTCACTGTCA TTAAAGATA ATGTG 945

55

(2) INFORMATION FOR SEQ ID NO: 169:

(i) SEQUENCE CHARACTERISTICS:  
60 (A) LENGTH: 902 base pairs  
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 169:

5  
GGCAGAGCCA CAGGAAGGAT GAGGAAGACC AGGCTCTGGG GGCTGCTGTG GATGCTCTTT 60  
GTCTCAGAAC TCCGAGCTGC AACTAAATTA ACTGAGGAAA AGTATGAACT GAAAGAGGGG 120  
10 CAGACCCTGG ATGTGAAATG TGA CTACTACG CTAGAGAAGT TTGCCAGCAG CCAGAAAGCT 180  
TGGCAGATAA TAAGGGACGG AGAGATGCCC AAGACCCTGG CATGCACAGA GAGGCCTTCA 240  
AAGAATTTCC ATCCAGTCCA AGTGGGGAGG ATCATACTAG AAGACTACCA TGATCATGGT 300  
15 TTACTGCGCG TCCGAATGGT CAACCTTCAA GTGGAAGATT CTGGACTGTA TCAGTGTGTG 360  
ATCTACCAGC CTCCCAAGGA GCCTCACATG CTGTTCCGATC GCATCCGCTT GGTGGTGACC 420  
20 AAGGGTTTTT CAGGGACCCC TGGCTCCAAT GAGAAATCTA CCCAGAATGT GTATAAGATT 480  
CCTCCTACCA CCACTAAGGC CTTGTGCCA CTCTATACCA GCCCCAGAAC TGTGACCCAA 540  
GCTCCACCCA AGTCAACTGC CGATGTCTCC ACTCCTGACT CTGAAATCAA CCTTACAAAT 600  
25 GTGACAGATA TCATCAGGT TCCGGTGTTC AACATTGTCA TTCTCCTGGC TGGTGGATTG 660  
CTGAGTAAGA GCCTGGTCTT CTCTGTCTG TTGTCTGCA CGCTGAGGTC ATTTGTACCC 720  
30 TAGGCCCCAG AACCCACGAG AATGTCTCT GACTTCCAGC CACATCCATC TGGCAGTTGT 780  
GCCAAGGGAG GAGGGAGGAG GTAAAAGGCA GGGAGTTAAT AACATGAATT AAATCTGTAA 840  
TCACCRGCTA AAAAAAAAAA AAAAAAACN CGANCCCTNGG TTTTCAGCTC CATCAGCTCC 900  
35 TT 902

40

(2) INFORMATION FOR SEQ ID NO: 170:

(i) SEQUENCE CHARACTERISTICS:

45

(A) LENGTH: 1883 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 170:

AGAAAACAAC TGAAAACCA CATTTTCTA CATACAGCTG GGGAGGTAGC TGAGAACTTG 60  
GCACTGCGCA CACATACTAG GTTGAAAGAG AGTTGAGGAA ACCAGAAGGC CAAGTGGATC 120  
55 TGCTGGCAAA CCCTGAACCT GTCTCCTGCG CTGTCTCTAC AGTTCTGAAG TTGAAAATCC 180  
TTTTCATGCC TAGCATCTGC TTGAGTTATA AACCCCAAGG CAGCCATGTC ATAGACTAGT 240  
60 GTTACTCTT GTTTTGACTT TGTTTTAATG CTTCTAAGA CCCAAGTGCC TCCTGCTGTT 300

	TCCTCCTTTG TGGTAGCCTC TGGCCATCTG GGACCTCAAT CCCCAGCTTT CCCACTTTCA	360
	GCAGTCCTTT GCTCTCTTTG CTTCTACCTC AAATAGCCCC AGGAGTGGGC TTTAGTCTCC	420
5	AATATGGAGC ATYTCAAGCT TCTCTGGGG GATGGGGATT GGGATGGGCA GAATCTGTTT	480
	TGGWTCCTCC GGTATTTTCC AGTGGGTGTA AAAGCAGAGC TGGGCCTTTC CCTCTCTTAT	540
10	CCCTGAGGGT GGGTAAGAAG GACTGTATCT ACACCTGTTT TTCCCTACCT TCTCTTTTGT	600
	TAGGGAGGCC TCATTCTAAG TTCCTCAAGA GAGTCCTTGG CTTAAAGCTG TAGCAAGGGT	660
	GTGCTAGGTG GGGGATTTGG AGCAAAACCG TCGAGTAGGC ATGATACTGG TATGGAGTGG	720
15	GCCTGCAAAA TCAGACAGAA ATGGCTTGAG AAGCCGCAGG GGAGCATGCC TGTCTCTCAG	780
	TGATAGAGTA TGGGAGGGAC CTCCTAGCT TGGAAAATGA GAATTGAAGG GGTATGAAC	840
20	AAATAGGATG CCTAGTTGAG GATGTTCCCA AAGTTTGTG CAATCTTATC ATTAGTAGAT	900
	TTTATAAGCC ACAGAGACAA ACCAGAAACG GAATAATGTT ACTTTGGATG CTTTATTTTT	960
	TGTCTTAGG TGTGGCTTTG TACATGCAGA AGAATGCTAT ATGCTGCACA TTTTGCCTTT	1020
25	AAAGTCTTAC GACTTTCCCC ATTTTAGTCT AATGGGAAGA TACAGATGTG CAAGTCTGCT	1080
	TTTTTGTTTT TTGTTATTAT TTTTTTTTTT TTGCTCTGTG TTATGGACAT TTTCAGACAT	1140
30	GCACAGAACT GGAGAGGATG GTCCTTGAC CCCATGTGTC CATCACCTAG CTGCATCACT	1200
	TATCAGCTAT GGTCACCTG GTTTCATCTG TATCTCTCTC TTTTCACCTG TATTGTTTAT	1260
	TGAAAAATCA AGACACTATG CCAATGCAAC CGTGACTION TTGGGAGATT GGTAGTCTCT	1320
35	TTTGATGGTG ATAGTGATGG GGTGCACTAT CATAATCACA TCAGGTCGTC TTTTGTCTTT	1380
	TAATGITAAC TAATGAAGTT CCAGAGATGG GCCTTAGAAA TGTGTTTTAA GAATTAACAA	1440
40	GGAGTCTCAA AAAGAAATGA GAGGGATGCT TCCTTTCCCC TTGCATCTAC AAAACAAGAG	1500
	AGAGACTGTT CTGTTGTAAA ACTCTTTCAA AAATTCTGAT ATGGTAAGGT ACTTGAGACC	1560
	CTTCACCAGA ATGTCAATCT TTTTCTCTGT GTAACATGGA AACTTGTGTG ACCATTAGCA	1620
45	TTGTTATCAG CTGTACTGCG TCTCATAACT CTGGTTTGG AAGAATAATT TGGAATTTGT	1680
	TGCTGTGTTT TGTGAAAATA ACCTCCCCAA AATAATTAGT AACTGGTTGT TCTACTTGGT	1740
50	AATTTGACAC CCTGTTAATA ACGCAATTAT TTCTGTGTTT TTAACAGTA TAAATAGTTG	1800
	TAAGTTTGCA TGCATGATGG AAAAAATAAA ACCTGTATCT CTGTTAAAAA AAAAAAAAAA	1860
	AAAAAAAAAA AAAAAAAAAA AAA	1883
55		

(2) INFORMATION FOR SEQ ID NO: 171:

60 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2100 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 171:

	TACTTTTAGA TTACTGCCT TCAAAAAGTG CCTATTCTGA GCAACATAAA CGTTATTCCT	60
10	TACATATGTA TGTACACACG GTACCCAGAG TCGTACTGTG GCAGCCTTCA AAAACATACC	120
	ATCAGAAAGA GTAGGTGCTG AGATAAGGNA ACTTTGCCAA ATGNAAGAAA GTCACCTCACT	180
15	TCCAATATCC CCTCTTCAAG CGGCTACCGT GRAASGGGCT GCAAACACAT TCCCTGAGCA	240
	TCCCTTGCTG ATACAGCTTC TTTATATTTA TATCCTACTG GATGGTAGCA TATTGCTAAG	300
	GTTTCCTGTA CTCTGCTTCA AGGGAATGTA AGYTTTATGG CATGAAACA TTTAGGAAAA	360
20	AAAAAGATGT TTAAGAGAAT TAATAGAGCC GTAGTCTGTA TTAGGATGTG TGTCATATGT	420
	GTGTCTATA AACTAAGCAT CGGTGGGTTT AGAGTGTTAA AGTGTACGCA CATTCCTTCT	480
25	CCTTTTGTCT CTCAGGCTAA CATGAGAGAA AATAGAAAAG TCTTGGCTGT GGGGATTGGA	540
	AGCTCAGGGG GCCAAATGTC CTTGCCAGAT CCTTAGAGCA TTACTTTGAC TCCTAAAAAT	600
	AGTAGTGTAT GTTATTTGAT GGCTTTTGTT TCCATAGTTC CATCACTGAC AAAACTGTCA	660
30	ATACTGTGTA TGGAGCAGCA GCATAGCCTA GAGTGATGCA TTCTTACCCA GAGGTGGCAA	720
	TAGGAGAGGG TCCATGTAAA TAGGACGAGG TAGACAGTGC ATGATTGTAG GAGAAGGGTT	780
35	GAAGGGAGGA CATGATTCCA AAAAAGATCG TTCTCAATGT GTCGTCTGAC TCAACCAGCT	840
	GGCAGATTAC ACTTGCCAAG TCGTTCCTTT TCCTTCTAAG TCAGTTGGCT CCATATTAC	900
	TTGAATATGC CTCTGTTTGG GCAAAGCAAG ATACCTCCAC TTAACCTTTA TCCAAGGAAG	960
40	CTCTTGGTGT CCTCTTGGTC ATAAAGTTGT CTCCTACCTA ACCCAGTTTT ACCAAATGGA	1020
	AGTAAAAGGG GACAACTAT GGAAGATGGA CTCCATGCCA TTGCAGTCAG CCACCATTCT	1080
45	CTTTTCCATA TAAGGAGCCC CATTACATAA GCTACGGTG AGGTGGAAC AGCTATGTTT	1140
	CATAATTTC AAGAGTGTGAC CACCCTGCTC TAGTCATCAT CATTTGGATGA ATCCAGTTGA	1200
	CTCTTTGGCA AAAGGGTGAT ACTTTTCACT AAAAATGCCT ACTCTTCCTG TTGATGTTCC	1260
50	TTTTCTGTTT TTACCTTGTC CAATTCCAC ACTAGTCATT TTTTTTATTT TTTAGAGGAT	1320
	CAGATTTTAG CGCTGGAAAA TGAGTTCAAA AATTTCACTG TAATGTCATA AGGATGTTGG	1380
55	GATACAGAGA TTTTTTTTTT CCTTGGAAAC AAATGGACTG GGAAGAAACA CAGCATGGCT	1440
	TTGCTCTGAG TTTCAATCTG ATGATTATGA CCATGGAAGA TAGTCTTATG TAAAGGTTAA	1500
	ATGGTGTTTA CAAGTGGATA GATAAGGCGG AGATGGTGAG AAGCCGGGTT TTCTCTATGC	1560
60	TAAATGTGTC TACTAAGAGC AGCACTTCCT ACTAGCTAAG CACAATCATA GCCCCACCGT	1620

	GATGAGCTGC TAGTCTGAAT AACATTCCT GACTTAGGGA AAGGCACACA AAAACATATA	1680
5	AAGAATATGT CTATTTTCAT ATGTGTGATA CTGACAGAGC CATGGTATTC CTAAAATATA	1740
	GGTTCTCTT TTTCTGTGA TTCTTAGCAA ATTGCATTTA TTCCTACAT TACAAACCAT	1800
	CACTGATGTA TCCAAATAG CACACATAGT TCAGTATGAA AATAAGAGAA TAAATCTGT	1860
10	TATAAGCAAG TGATTTAGGT ATTTCTTTT GTGTTTATGC ATTATCTGAC TATATTAAAA	1920
	CCTGTTTTTC TATTTACCTT CTATCAGTTT TCTCTACCAA TTATGTTTTT TCAATGCTCT	1980
15	ATAAGAATGA ATATGGAAAT TATATTTCTT TTTCTGTAA AAGAGTTGCA ACTACTTTAT	2040
	TATATTTAGA AATCCAATAA ACTTCTTATT ACATTTAAAA AAAAAAAAAA AAAACTCGAA	2100
20	(2) INFORMATION FOR SEQ ID NO: 172:	
	(i) SEQUENCE CHARACTERISTICS:	
25	(A) LENGTH: 1930 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
30	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 172:	
	CCTTTGANTG TGGTCCCGGG TGCNGATTGG CAGCGCCTCC GCCGCGGCTC GTGGTTGTCC	60
	CGCCATGGCA CTGTGCGGGG GGCTGCCCGG GGAGCTGGCT GAGGCGGTGG CCGGGGGCCG	120
35	GGTGCTGGTG GTGGGGGCGG GCGGCATCGG CTGCGAGCTC CTCAAGAATC TCGTGCTCAC	180
	CGGTTCTCC CACATCGACC TGATTGATCT GGATACTATT GATGTAAGCA ACCTCAACAG	240
40	ACAGTTTTTG TTTCAAAAGA AACATGTTGG AAGATCAAAG GCACAGGTG CCAAGGAAAG	300
	TGTACTGCAG TTTTACCCGA AAGCTAATAT CGTGCCTAC CATGACAGCA TCATGAACCC	360
	TGACTATAAT GTGAATTTT TCCGACAGTT TATACTGGTT ATGAATGCTT TAGATAACAG	420
45	AGCTGCCCGA AACCATGTTA ATAGAATGTG CCTGGCAGCT GATGTTCTTC TTATTGAAAG	480
	TGGAACAGCT GGGTATCTTG GACAAGTAAC TACTATCAAA AAGGTGTGA CCGAGTGTTA	540
50	TGAGTGTGAT CCTAAGCCGA CCCAGAGAAC CTTTCCTGGC TGTACAATTC GTAACACACC	600
	TTCAGAACCT ATACATTGCA TCGTTTGGGC AAAGTACTTG TTCAACCACT TGTPTGGGGA	660
	AGAAGATGCT GATCAAGAAG TATCTCCTGA CAGAGCTGAC CCTGAAGCTG CCTGGGAACC	720
55	AACGGAAGCC GAAGCCAGAG CTAGAGCATC TAATGAAGAT GGTGACATTA AACGTATTTC	780
	TACTAAGGAA TGGGCTAAAT CAACTGGATA TGATCCAGTT AAACCTTTTA CCAAGCTTTT	840
60	TAAAGATGAC ATCAGGTATC TGTGACAAT GGACAACTA TGGCGGAAAA GGAAACCTCC	900

AGTTCGGTTG GACTGGGCTG AAGTACAAAG TCAAGGAGAA GAAACGAATG CATCAGATCA 960  
ACAGAATGAA CCCAGTTAG GCCTGAAAGA CCAGCAGGTT CTAGATGTAA AGAGCTATGC 1020  
5 ACGTCTTTTT TCAAAGAGCA TCGAGACTTT GAGAGTTCAT TTAGCAGAAA AGGGGGATGG 1080  
AGCTGAGCTC ATATGGGATA AGGATGACCC ATCTGCAATG GATTTTGTC A CCTCTGCTGC 1140  
10 AAACCTCAGG ATGCATATTT TCAGTATGAA TATGAAGAGT AGATTGATA TCAAATCAAT 1200  
GGCAGGGAAC ATTATTCCTG CTATTGCTAC TACTAATGCA GTAATTGCTG GGTTGATAGT 1260  
ATTGGAAGGA TTGAAGATTT TATCAGGAAA AATAGACCAG TGCAGAACAA TTTTTTGTAA 1320  
15 TAAACAACCA AACCCAAGAA AGAAGCTTCT TGTGCCTTGT GCACTGGATC CTCCCAACCC 1380  
CAATTGTTAT GTATGTGCCA GCAAGCCAGA GGTGACTGTG CGGCTGAATG TCCATAAAGT 1440  
GACTGTTCTC ACCTTACAAG ACAAGATAGT GAAAGAAAAA TTTGCTATGG TAGCACCAGA 1500  
20 TGTCCAAATT GAAGATGGGA AAGGAACAAT CCTAATATCT TCCGAAGAGG GAGAGACGGA 1560  
AGCTAATAAT CACAAGAAGT TGTGAGAATT TGAATTAGA AATGGCAGCC GGCTTCAAGC 1620  
25 AGATGACTTC CTCCAGGACT ATACTTTTATT GATCAACATC CTTCATAGTG AAGACCTAGG 1680  
AAAGGACGTT GAATTGTAAG TTGTTGGTGA TGCCCCGGAA AAAGTGGGGS CCAAACAAGC 1740  
TGAAGATGCT GCCAAAAGCA TAACCAATGG GCAGTGATGA TGGGAGCTTC AGCCCTCCAC 1800  
30 CTYCACAGCT TCAAGGAGGC AAGATGGACG TYTCYCATAG TTGATYCGGR TGAAGAAGRT 1860  
TCTCCAATAA TTGCCCGACG TTCATTGAAG GAAGGAGGAG GAGGCCCGCC AAGAGGGGAA 1920  
35 TTTAGGNTTG 1930

40 (2) INFORMATION FOR SEQ ID NO: 173:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1509 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 173:

50 GGCCCTGGCC TCTGGGCTGA GGCTTGCTAG GGACTCGGGG TGGCTCTAAG GGGCAGGGAT 60  
AGGGCTGGGG AGCGCCGGCC TGTGGCCCTG ACCAGCCCCT TCTCGTGCRG GTTCCACCCC 120  
GATGCAGGTG GTACAGTGCT TGACGCGGGA CAGCTACCTG ACGCACTGCT TCCTCCAGCA 180  
55 CCTCATGGTC GTGCTGTCTT CTCTGGAACG CACGCCCTCG CCGGAGCCTG TTGACAAGGA 240  
CTTCTACTCC GAGTTTGGGA ACAAGACCAC AGGAAGATG GAGAACTACG AGCTGATCCA 300  
60 CTCTAGTCGC GTCAAGTTTA CCTACCCAG TGAGGAGGAG ATTGGGGACC TGACGTTTAC 360

	TGTGGCCCAA AAGATGGCTG AGCCAGAGAA GGCCCCAGCC CTCAGCATCC TGCTGTACGT	420
5	GCAGGCCCTTC CAGGTGGGCA TGCCACCCCC TGGGTGCTGC AGGGGCCCCC TGCGCCCCAA	480
	GACACTCCTG CTCACCAGCT CCGAGATCTT OCTCCTGGAT GAGGACTGTG TCCACTACCC	540
	ACTGCCCGAG TTTGCCAAAG AGCCGCCGCA GAGAGACAGG TACCGGCTGG ACGATGGCCG	600
10	CCGCGTCCGG GACCTGGACC GAGTGTCTAT GGGCTACCAG ACCTACCCGC AGCCCTCACC	660
	CTCGTCTTCG ATGACGTGCA AGGTCATGAC CTCATGGGCA GTGTACCCCT GGACCACTTT	720
15	GGGAGGTGC CAGGTGGCCC GGCTAGAGCC AGCCAGGGCC GTGAAGTCCA GTGGCAGGTG	780
	TTTGTCCCA GTGCTGAGAG CAGAGAGAAG CTCATCTCGC TGTGGCTCG CCAGTGGGAG	840
	GCCCTGTGTG GCCGTGAGCT GCCTGTGAG CTCACCGGCT AGCCAGGGCC ACAGCCAGCC	900
20	TGTCGTGTCC AGCCTGACGC CTACTGGGGC AGGGCAGCAG GCTTTTGTGT TCTCTAAAAA	960
	TGTTTTATCC TCCCTTTGGT ACCTTAATTT GACTGTCTC GCAGAGAATG TGAACATGTG	1020
25	TGTGTGTGT GTTAATTCCT TCTCATGTTG GGAGTGAGAA TGCCGGGCC CTCAGGCTG	1080
	TCGGTGTGCT GTCAGCCTCC CACAGGTGGT ACAGCCGTGC ACACCACTGT CGTGTCTGCT	1140
	GTGTGGGAC CGTGTTAAC ACGTGACACT GTGGGTCTGA CTTTCTCTTC TACACGTCCT	1200
30	TTCTGAAGT GTCGAGTCCA GTCCTTTGTT GCTGTGCTG TTGCTGTGC TGTGCTGTT	1260
	GGCATCTGCG TGCTAATCCT GAGGCTGGTA GCAGAATGCA CATTGGAAGC TCCCACCCCA	1320
35	TATTGTCTT CAAAGTGGAG GTCTCCCTG ATCCAGACAA GTGGGAGAGC CCGTGGGGC	1380
	AGGGGACCTG GAGCTGCCAG CACCAAGCGT GATTCTGCT GCCTGTATTC TCTATTCAA	1440
	TAAAGCAGAG TTTGACACCG TCAAAAAA AAAA AAAA ATTNCTGCGG	1500
40	CCTCAAGGG	1509

45 (2) INFORMATION FOR SEQ ID NO: 174:

(i) SEQUENCE CHARACTERISTICS:

- 50 (A) LENGTH: 3173 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 174:

55	TGACCCCA GCGTCCGTGC TTTTCCACAG AAGGTTAGAC CCTGAAAGAG ATGGCTCAGC	60
	ACCACCTATG GATCTGTGTC CTTTGCCTGC AAACCTGGCC GGAAGCAGCT GGAAAAGACT	120
60	CAGAAATCTT CACAGTGAAT GGGATTCTGG GAGAGTCAGT CACTTTCCT GTAAATATCC	180



	AAGAACCACG GCAAGTTAAA ATCATTGCTT GGAATTCTAA AACATCTGTT GCTTATGTAA	240
	CACCAGGAGA CTCAGAAACA GCACCCGTAG TTAAGTGTAC CCACAGAAAT TATTATGAAC	300
5	GGATACATGC CTTAGGTCCG AACTACAATC TGGTCATTAG CGATCTGAGG ATGGAAGACG	360
	CAGGAGACTA CAAAGCAGAC ATAAATACAC AGGCTGATCC CTACACCACC ACCAAGCGCT	420
10	ACAACTGCA AATCTATCGT CGGCTTGGGA AACCAAAAAT TACACAGAGT TTAATGGCAT	480
	CTGTGAACAG CACCTGTAAAT GTCACACTGA CATGCTCTGT AGAGAAAGAA GAAAAGAATG	540
	TGACATACAA TTGGAGTCCC CTGGGAGAAG AGGGTAATGT CCTTCAAATC TTCCAGACTC	600
15	CTGAGGACCA AGAGCTGACT TACACGTGTA CAGCCAGAA CCCTGTCAGC AACAATTCTG	660
	ACTCCATCTC TGCCCGGCAG CTCTGTGAGC ACATGCAAT GGGCTTCCGT ACTCACCACA	720
20	CCGGTTGCT GAGCGTGTG GCTATGTTCT TTCTGCTTGT TCTCATCTG TCTTCAGTGT	780
	TTTGTTCGG TTTGTTCAG AGAAGACAAG ATGCTGCCTC AAAGAAAACC ATATACACAT	840
	ATATCATGGC TTCAAGGAAC ACCCAGCCAG CAGAGTCCAG AATCTATGAT GAAATCCTGC	900
25	AGTCCAAGGT GCTTCCCTCC AAGGAAGAGC CAGTGAACAC AGTTTATTCG GAAGTGCAGT	960
	TTGCTGATAA GATGGGAAA GCCAGCACAC AGGACAGTAA ACCTCCTGGG ACTTCAAGCT	1020
30	ATGAAATGT GATCTAGGCT GCTGGGCTGA ATTCTCCCTC TGGAACTGA GTTACAACCA	1080
	CCAATACTGG CAGGTTCCCT GGATCCAGAT CTCTCTGCC CAACTCTTAC TGGGAGATTG	1140
	CAAAGTCCA CATCTCAGCC TGTAAAGCAA GCAGGAAACC TTCTGCTGGG CATAGCTTGT	1200
35	GCCTAAATGG ACAAATGGAT GCATACCTT CCTGAAATGA CTCCCTTCTG AATGAATGAC	1260
	AAAGCAGGTT ACCTAGTATA GTTTTCCCAA ACTTCTTCCC ATCATAGCAC ATGTAGAAAA	1320
40	TAATATTTTT ATGGCACACT GGGATAAACA AGCAAGATTG CTCACTTCTG GAAGCTGCAT	1380
	ATGACTAGAG GCCTCTTGTG ACTGGAGGTA ACAACCTGC CCAGTAACTG TGGGAGAAGG	1440
	GGATCAATAT TTGACACACC TGTAAATAGG CATGGCACAC CAGCCAAGAT GCTCTGCTCA	1500
45	CAGTCAGTAT GTGTGAAGAT CCTGGTGGC TGGCCTTCC CAGCATCTT GAGCAAATTA	1560
	GGAAAATGTA CCCTTCGCTT GAGGCAGATG CAGCCCTCC CCCGAGTGA TGGCTTGGAG	1620
50	AGCAGAATGT GGGCTGCATA TAAGCACACT CATCCCTTTG TCTGGGAATC TTTGTGCAGG	1680
	GCATAACAGG CTTAGTAAGT CCAAACACAG ATGACAGTGC TGTGTGGTC TCTGTCAGAG	1740
	TTGTGGCTCT CAGCCATGTA GACACACTCT CCAAATGGAG TGTGGAAAA TGTCTTTCT	1800
55	GCAGGTCTA GAGACTGCTG GGACACTTTT CTGGAGTGC TACTTCAGAA GCCTTATAGG	1860
	ATTTTCTTTC TGGCCAAGAT TTCTTCTGT ATCACTCCAA GCAGCCTCAG CAGAAGAAGC	1920
60	AGCCATGCCC AGTATTCCCA CTCTCCAAA GGAAGTACC AGCTTATATT TCTCACACTT	1980

	CTGGGGAAC TGGTATAATC CAACCATCAA AATAGAAGAC CTTGCAAGAA GCAGAGTCAT	2040
	TCTCCAGAAG GAACTTGGGA GATGATGGTG CAGATGATGA AACTGGGTTC ATCCCAGTTC	2100
5	CAAAGACTCA GAGAACTAGA GTTTAAGCTG AGGCAGAGTG CCGCCACCCT GGCATGCCCC	2160
	ACAAACAGAT CACCAGCCAG CTTACACAGG CATTAACTCT CCTCAATGAG GAAGAATCAT	2220
10	TCACAACTGA GCAAGACATT CATATGATCA TTTAAGGAAG TGTTTCCCTT ATGTGTTAGC	2280
	AAGTATAATC GGCTAACTCC TAAATCCCAA TGAATAGTCC TAGGCTGGAC AGCAATGGGC	2340
	TGCAATTAGG CAGATAAAGA CATCAGTCCC AGTAAATGAA TCCATAGACT CATCTAGCAC	2400
15	CAACTACCAT TAGCACTATG TTAGGAGCTG CAAGGCCCCA AAGTAGAAGA TGTGCATAAT	2460
	GTCTGCTCTT GTGTAGCTCA GGAGACAATT CCAGCACAGA CACTACAGTT AACGCTGAAC	2520
	TGCAGCTGCA AGTAATAGCA TGAACAGTCA GAAAAATACC TTATGAGGGG GCAGGGCTGA	2580
20	AGCTGGGCCT TGAAGGATGG ATGAAATTTG GATAGAGAAT GAGGAAGACA GAGGGCCTCC	2640
	AAGTGAGAGA AGCATGAAAA ATGAGCAGGG GCCTGGATCA GTGGGTGTA TTCAGAGCAC	2700
25	CTCTCCAGAT GCACCATGCA TGCTCACAGT CCCTTGCCTA TGTGTGGCAG AGTGTCCCAG	2760
	CCAGATGTGT GCCCCCACC CATGTCCATT TACATGTCTT TCAATGCCCA CCTCAAAAGG	2820
30	TACCTCTTCT GTAAAGCTTT CCTGGTATC AGGAATCAAA ATTAATCAGG GATCTTTTCA	2880
	CACTGCTGTT TTTTCTCTT TGGTCCTTCT ATCACTAAAA CTCATCTCAT TCAGCCTTAC	2940
	AGCATAACTA ATTATTGTGT TTCTCTACTA CATTGTACAT GTGGGAATTA CAGATAAACG	3000
35	GAAGCKGCT GGGGTGGTGG CTCACGCTG TAATCCCAAC ACTTTGGGAG GCCAAGGCAG	3060
	GCGGATCACC TGAGGTCAGG ARTTCGAGAT TARTCTGGCC AACATGGTGA AACCCCATNT	3120
40	NTACTAAAAA TACGAAATTA GCCAGGTGTG GTGGCACACA TCTGTAGTCC CAG	3173

45 (2) INFORMATION FOR SEQ ID NO: 175:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 991 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 175:

55	AAATTCGGCA CAGCTGAGAG GAGACACAAG GAGCAGCCCG CAAGCACCAA GTGAGAGGCA	60
	TGAAGTTACA GTGTGTTTCC CTTTGGCTCC TGGGTACAAT ACTGATATTG TGCTCAGTAG	120
	ACAACCACGG TCTCAGGAGA TGTCTGATTT CCACAGACAT GCACCATATA GAAGAGAGTT	180
60	TCCAAGAAAT CAAAAGAGCC ATCCAAGCTA AGGACACCTT CCCAAATGTC ACTATCCTGT	240

	CCACATTGGA GACTCTGCAG ATCATTAAGC CCTTAGATGT GTGCTGCGTG ACCAAGAACC	300
5	TCCTGCGGTT CTACGTGGAC AGGGTGTTC AAGATCATCA GGAGCCAAAC CCCAAAATCT	360
	TGAGAAAAAT CAGCAGCATT GCCAACTCTT TCCTCTACAT GCAGAAAACT CTGCGGCAAT	420
	GTCAGGAACA GAGGCAGTGT CACTGCAGGC AGGAAGCCAC CAATGCCACC AGAGTCATCC	480
10	ATGACAACTA TGATCAGCTG GAGGTCCACG CTGCTGCCAT TAAATCCCTG GGAGAGCTCG	540
	ACGTCTTTCT AGCCTGGATT AATAAGAATC ATGAAGTAAT GTCCTCAGCT TGATGACAAG	600
15	GAACCTGTAT AGTGATCCAG GGATGAACAC CCCCTGTGCG GTTTACTGTG GGAGACAGCC	660
	CACCTTGAAG GGAAGGAGA TGGGAAGGC CCCTTGCAGC TGAAAGTCCC ACTGGCTGGC	720
	CTCAGGCTGT CTTATTCCGC TTGAAAATAG CAAAAAGTC TACTGTGGTA TTTGTAATAA	780
20	ACTCTATCTG CTGAAAGGGC CTGCAGGCCA TCCTGGGAGT AAAGGGCTGC CTTCCCATCT	840
	AATTTATTGT GAAGTCATAT AGTCCATGTC TGTGATGTGA GCCAAGTGAT ATCCTGTAGT	900
25	ACACATTGTA CTGAGTGGTT TTTCTGAATA AATTCCATAT TTTACCTAAA AAAAAAAAAA	960
	AAAAACTCGA GGGGGGGCCC GTACCCAATT T	991

30

(2) INFORMATION FOR SEQ ID NO: 176:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1290 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

35

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 176:

	ACAGCCCTCT TCGGAGCCTG AGCCCGGCTC TCCTCACTCA CCTCAACCCC CAGGCGGCC	60
	CTCCACAGGG CCCCTCTCCT GCTGGACGG CTCTGCTGGT CTCCCGTCC CCTGGAGAAG	120
45	AACAAGGCCA TGGGTGGCC CTTGCTGCTG CCCCTCTGCT YCCTGCTGCTW GCGCCAGCA	180
	TTTCTGCAGC CTRGTGGCTC CACAGGATCT GGTCCAAGCT ACCTTTATGG GGTCACTCAA	240
50	CCAAAACACC TCTCAGCCTC CATGGGTGGC TCTGTGGAAA TCCCCTTCTC CTTCTATTAC	300
	CCCTGGGAGT TAGCCAYAGY TCCCRACGTG AGAATATCCT GGAGACGGGG CCACCTCCAC	360
	GGGCACTCCT TCTACAGCAC AAGGCCGCTT TCCATTACCA AGGATTATGT GAACCGGCTC	420
55	TTTCTGAACT GGACAGAGGG TCAGGAGAGC GGCTTCCTCA GGATCTCAA CTTGCGGAAG	480
	GAGGACCACT CTGTGTATTT CTGCCGAGTC GAGCTGGACA CCCGAGATC AGGGAGGCAG	540
60	CAGTTGCAGT CCATCAAGGG GACCAAATC ACCATCACCC AGGCTGTCAC AACCACCACC	600

ACCTGGAGGC CCAGCAGCAC AACCACCATA GCCGGCCTCA GGGTCACAGA AAGCAAAGGG 660  
 CACTCAGAAT CATGGCACCT AAGTCTGGAC ACTGCCATCA GGGTTGCATT GGCTGTCGCT 720  
 5 GTGCTCAAAA CTGTCATTTT GGGACTGCTG TGCCTCCTCC TCTGTGGTGG AGGAGAAGGA 780  
 AAGGTAGCAG GCGGCAAGC AGTGACTTCT GACCAACAGA GTGTGGGGAG AAGGGATGTG 840  
 10 TATTAGCCCC GGAGGACGTG ATGTGAGACC CGCTTGTGAG TCCTCCACAC TCGTTCCCA 900  
 TTGGCAAGAT ACATGGAGAG CACCCTGAGG ACCTTTAAAA GGCAAAGCCG CAAGGCAGAA 960  
 GGAGGCTGGG TCCCTGAATC ACCGACTGGA GGAGAGTTAC CTACAAGAGC CTTTCATCCAG 1020  
 15 GAGCATCCAC ACTGCAATGA TATAGGAATG AGGTCTGAAC TCCACTGAAT TAAACCACTG 1080  
 GCATTTGGGG GCTGTTYATT ATAGCAGTGC AAAGAGTTCC TTTATCCTCC CCAAGGATGG 1140  
 AAAATACAAT TTATTTTGCT TACCATACAC CCTTTTCTC CTCGTCCACA TTTTCCAATC 1200  
 20 TGTATGGTGG CTGCTTCTA TGGCAGAAGG TTTTGGGGAA TAAATAGCGT GANATGMTC 1260  
 TGACTNAAAA AAAAAAAAAA AAAAATCGA 1290

25

## (2) INFORMATION FOR SEQ ID NO: 177:

30

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2290 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

35

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 177:

TGGGGCCCCCT TTTGGATGCT CTGGGTGTTT TTGCCAAGAG TTACAGGATG TCAAGTGTGG 60  
 40 GGAGCTCAGC ACCCTTGCTG TGGACCAAGT AAGGCTGTTT CAGACCAGGT GCTTCCAGAC 120  
 ATTTCCAGGC TCCAGGAGAG AGGCTGGGAG CCCCCACAGA AAGCACAGGA AAATGCAAAA 180  
 45 AAAAAACAGT CTTTITTTTT TTTTGTCTTT TTATTATGAA AACAAAACAA ATGCCCCAGG 240  
 AGAAGGGTCC ATGATTACCA GAAACATCAA AGAGTACTTT CTACCATTTT TATTCTGTG 300  
 TGTGAGGCC AGCATTGCAA TAAACAAGCT AACTACTTA CATTGGACTC ATTTTCAGTA 360  
 50 ACTGACATTT ACAGGAATAT ACTAGAAACG GACTAAAAA GTTTAAGAAA AGTTACGGTA 420  
 AACTTGCATG CACATCATAC AGAAAAGTAA CATTTTAAAT ATAAAAAGA AAAACTTCCT 480  
 55 GGAAGCATTG TGCCAGTATT AAGGAACAGT GCTACTCTGG ATGTGACAAA TTCTGTATGT 540  
 GGGTGTTACT CTTTCCCAAA AGACTGTCAG AGGCGTGAGT GCTGCAAAAG AACACAACA 600  
 AAAACAACA CAAAAAAA TGTGCTTAC AGTTTGTAAAG CAAGATGACA CTGCCCAACA 660  
 60 CAAAGAGGGG TCTGGAGTTC AGTTACGCC CGAAGCCTGC CCCCTCGGCC TCCAGGGGTC 720

	ATTCAGAGTG TTCTCAAATC CAATTCCGAC ACACGACTTG TCACTACTCC TCTCCCCTTG	780
5	AAAAAAGCAT GTTAGAAGCT GCCCTACAGG TCTCAGCAGT GGGACAATCT AATTGAATCA	840
	CCGCAGCCTT CTAATACAGA AGAAACGGAC GTGACTGTCA CCCTCAGCCC GCCAGCAAGG	900
	GCGCTGAGGA AGTCATTAAT CCTTCGAAAC TCTGAAAAGA AACCAGTGTT GAAGTCTGGA	960
10	CAGAAAGCCT TAAAAAAGTG ACAGCACCAA TGCAGCTGCT CAGTGTACCC NCCGTGGGCT	1020
	GTCAGGTGCA GTGGCTTCTT TCTAGATGAA AGGAGCAGAG GCGAGCCGAC GCCACCGTCA	1080
15	CAGAGAACCA GCCGAGAAGG AAAGGCCCCA CGATGCTCCC TGTGCGCTGC CCCACAGCC	1140
	GGCCGCTCCC CCGACGGCTC ACACAGGCAG CACCTCACTG CCCTGTGGCT GGAGGGGCAT	1200
	TGCAAGGAGC GCCCCCAGC CCCAGGCACC CCGGCTTAG GGTGTACGTA TCACCCAGCC	1260
20	CTGTGCTGGC AGCAGCTTAC CAACCAGCCT GCGTGAAGAC CTGTCAACTG TCGTGTGTGA	1320
	ATTCCTTAAA TTCGGTTTAA ATAGTCCATT AAAGATCTGT TTAGAAAATA CCTTTGAAAA	1380
25	CGAGGGTAAC TTTAAAAAAT GGAAACTTTC AAATCCATTT ATATTTTAT TATAAACAA	1440
	ACTTAATTAA AAGTTTAACA AACTGGCTGA AAATCACCA AGTGTGAGAC TCACCAGCAA	1500
	TTTAAAAAAT GATAATTTAC CAGCATCTCC TCATCAGAGT TCCCTCTCCA GTAAGGGTAT	1560
30	ACCTACATCT GTAAGGGTCA GTGGACTCTG AATCAATTTT ATGGTTGTTT TAAATCACC	1620
	GTGTATTAGG ATACTAATGA TAGTCCCTAT ATCCATCCAG AAATGCTGGC AGAAAGCACT	1680
35	GGCCACCATA CAGGACAGAC CACACCACAG CTCCATACCC AGCGTCTGCC TGGAGGCTCC	1740
	CCCACGCTGA GGTCCGGGAG AATGCCTGGT TTCAGTCATT TCCGACTAA CTGTGACAAC	1800
	GCGTGAGCAG GGAGCACCGT GCGAGTCTCC GGGAGGGAAT CCTCCTGGGG CCCAGAGACT	1860
40	CCTCCACCCC TGGGGAGGGC AGACAGGCTC GGGARGGCCT GGCCAGGCCA CTGGAGGCTG	1920
	GCAGGGAGCA GGCATGTCCA CCCGCAAGCC TGGGAGGCTA ACTCTGGCAT TCCTGGCCCG	1980
45	AGCCGCCATG CTCATGGTG GGCCAGTTTG GGACATCCCC GTACTCAAAG ACCATATGGC	2040
	AGCCTCTGGG AAAACAAAAC CAAAACATCA CCTTCTATTA AACTCTGTAT ATTATTATTT	2100
	TTTACAATAG AAAGTTAAAA ATCAAGACTT AGATTTACTA TACATTTTTT CTCTCAGATT	2160
50	ACAAAGTTTA TATTATATAA CTGGGGTICC CTAAATTGAT TTCTTTTAAA ACAGTCTTAA	2220
	AGAGACCAGA AGTGAATACA AAAGAACTAA ACAAATAAAA AAATTAGAAT GTGCTGTAGC	2280
55	TGAAAGCTGT	2290

(2) INFORMATION FOR SEQ ID NO: 178:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 549 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 178:

GGCACGAGCC ATGCCTGGCC TCTCCTTGAT TCCTACAGTC ACTTGTGTGG CTGTTTCTGA 60  
CTCAGCAGCT ACCTGCATTG TGGCCAAAGG ATGACCTATT CCTTCTCAGG AGGGCAAAAA 120  
TGTGGAATAG TGTCTGTCCA TGCCTCTCCT CATGGGCTAC CACCTCTGCC ACCGTGGTTA 180  
ATCAGTAACA ACCAGGAGAG AAGCTGCTGG AACTGACCTC TGGGAACTCC CTGGGATGGT 240  
TTGGTGCAGG AATGTAGTAG GCATACACGT GGTTCGCTGG ATCTGGGCCC TCCTGATGTG 300  
AGTAGAGAGG TAAAAGGCCA CCATCTCCTT GACCTCTGGG GAACTCATCC ACAAGAAGA 360  
TGTTTCCAAG ATGCTTCTGA AGATTCCTTA AAAATAGCCG GTTTCACCC CCGTGAATGC 420  
ATCCATTCTA GAATGCTCCT TCACCAGGAC CAGAGAACTG ATTTACAGAA GTGACATGAA 480  
AACATTCCAT CCCAGAATTT GCAGTAGCTC AAATTAAGTT TCTAGCTATT AAAAAGAAAA 540  
AAAAA 549

## (2) INFORMATION FOR SEQ ID NO: 179:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1509 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 179:

GGCACGAGGG CTCATTCAATT CCGCGCCGGG CCTGCCAGAC ACCTGCGCCC TTCTGCAGCC 60  
GCCCCCGCA TCCGCCCGG CAGCCCCAG CATGTCCGGC CCAGACGTCG AGACGCCGTC 120  
CGCCATCCAG ATCTGCCGGA TCATGCGGCC AGATGATGCC AACGTGGCCG GCAATGTCCA 180  
CGGGGGGACC ATCCTGAAGA TGATCGAGGA GGCAGGCGCC ATCATCAGCA CCCGGCATTG 240  
CAACAGCCAG AACGGGGAGC GCTGTGTGGC CGCCTGGCT CGTGTGAGC GCACCGACTT 300  
CCTGTCTCCC ATGTGCATCG GTGAGGTGGC GCATGTCAGC GCGGAGATCA CCTACACCTC 360  
CAAGCACTCT GTGAGGTGC AGGTCAACGT GATGTCCGAA AACATCCTCA CAGGTGCCAA 420  
AAAGCTGACC AATAAGGCCA CCCTGTGTA TGTGCCCCTG TCGCTGAAGA ATGTGGACAA 480  
GGTCTCGAG GTGCCTCTG TTGTGTATTC CCGGCANGAG CAGGAGGAGG AGGGCCGGAA 540  
GCGGTATGAA GCCCAGAAGC TGGAGCGCAT GGAGACCAAG TGGAGGAACG GGGACATCGT 600

	CCAGCCAGTC CTCAACCCAG AGCCGAACAC TGTCAGCTAC AGCCAGTCCA GCTTGATCCA	660
5	CCTGGTGGGG CCTTCAGACT GCACCCCTGCA CGGCTTTGTG CACGGAGGTG TGACCATGAA	720
	GCTCATGGAT GAGGTGCGCG GGATCGTGGC TGCACGCCAC TGCAAGACCA ACATCGTCAC	780
	AGCTTCCGTG GACGCCATTA ATTTTCATGA CAAGATCAGA AAAGGCTGCG TCATCACCAT	840
10	CTCGGGACGC ATGACCTTCA CGAGCAATAA GTCCATGGAG ATCGAGGTGT TGGTGGACGC	900
	CGACCCGTGT GTGGACAGCT CTCAGAAGCG CTACCGGGCC GCCAGTGCCT TCTTCACCTA	960
15	CGTGTGCTG AGCCAGGAAG GCAGGTGCGT GCGTGTGCCC CAGCTGGTGC CCGAGACCGA	1020
	GGACGAGAAG AAGCGCTTTG AGGAAGGCAA AGGGCGGTAC CTGCAGATGA AGGCGAAGCR	1080
	ACAGGGCCAC GCGGASCYTC AGCCCTAGAC TCCCTCCTCC TGCCACTGGT GCCTCGAGTA	1140
20	GCCATGGCAA CGGGCCCACT GTCCAGTCAC TTAGAAGTTC CCCCCTTGGC CAAAAACCCA	1200
	ATTCACATG AGAGCTGGTG TTGTCTGAAG TTTCGTATC ACAGTGTTAA CCTGTACTCT	1260
25	CTCCTGCAAA CCTACACACC AAAGCTTTAT TTATATCAIT CCAGTATCAA TGCTACACAG	1320
	TGTTGTCCCG AGCGCCGGA GCGTGTGGC AGAAACCCTC GGAATGCTT CCGAGCACGC	1380
	TGTAGGTAT GGAAGAACC CAGCACCCT AATAAGCTG CTGCTTGGCT GGAAAAAAA	1440
30	AAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA	1500
	AGAAAAAAN	1509

35

(2) INFORMATION FOR SEQ ID NO: 180:

- 40 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1316 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

- 45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 180:

	AGCTGTATCA TAGGAAAGAT GGCCACACCG GCGGTACCAG TAAGTGCTCC TCCGGCCACG	60
50	CCAACCCAG TCCCGCGGC GGGCCAGCC TCAGTTCCAG CGCCAACGCC AGCACCGGCT	120
	GCGGCTCCGG TTCCCGCTGC GGCTCCAGCC TGCATCTCA GACCCGCGG CAGCAGCGC	180
	TGCAACTGCG GCTCCTGGCC AGACCCGCG CTCAGCGCAA NTCAGCGCA GACCCAGCG	240
55	CCGCTCTGC CTGGTCTGC TCTTCCAGG CCCTTCCCG GCGGCGCGT GGTGAGGCTG	300
	CACCCAGTCA TTTTGGCCTC CATGTGGAC AGCTACGAGA GACGCAACGA GGGTCTGCC	360
60	CGAGTTATCG GGACCTGTT GGAAGTGTG GACAAACACT CAGTGGAGGT CACCAATTGC	420

	TTTTCACTGC CGCACAAATGA GTCAGAAGAT GAAGTGGCTG TTGACATGGA ATTTGCTAAG	480
	AATATGTATG AACTGCATAA AAAAGTTTCT CCAAATGAGC TCATCCTGGG CTGGTACGCT	540
5	ACGGGCCATG ACATCACAGA GCACTCTGTG CTGNATCCAT GAGTACTACA GCGAGAGGC	600
	CCCCAACCCC ATCCACCTCA CTGTGGACAC AAGTCTCCAG AACGSCCGCA TGAGCATCAA	660
10	AGCCTACGTC AGCACTTTAA TGGGAGTCCC TGGGAGGACC ATGGGAGTGA TGTTCACGCC	720
	TCTGACAGTG AAATACGCGT ACTACGACAC TGAACGCATC GGAGTTGACC TGATCATGAA	780
	GACCTGCTTT AGCCCCAACA GAGTGATTGG ACTCTCAAGT GACTTGCAGC AAGTAGGAGG	840
15	GGCATCAGCT CGCATCCAGG ATGCCCTGAG TACAGTGTG CAATATGCAG AGGATGTACT	900
	GTCTGGAAAG GTGTCAGCTG ACAATACTGT GGGCCGCTTC CTGATGAGCC TGGTTAACCA	960
20	AGTACCGAAA ATAGTTCCCG ATGACTTTGA GACCATGCTC AACAGCAACA TCAATGACCT	1020
	TTTGATGGTG ACCTACCTGG CCAACCTCAC ACAGTCACAG ATTGCACTCA ATGAAAAACT	1080
	TGTAAACCTG TGAATGGACC CCAAGCAGTA CACTTGCTGG TCTAGGTATT AACCCAGGA	1140
25	CTCAGAAGTG AAGGAGAAAT GGGTTTTTTG TGGTCTTGAG TCACACTGAG ATAGTCAGTT	1200
	GTGTGTGACT CTAATAAACG GAGCCTACCT TTTGTAAATT AAAAAAAAAA AAAAAAACCN	1260
30	SGRGGGGGGG CCCGGTCCCA TTSSCCCTTT NGTAATTCGT NITACAATCC CCNGGC	1316

35 (2) INFORMATION FOR SEQ ID NO: 181:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 777 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 181:

45	GGCATGWKCA GACATGACTT CTATTGCCAG GCTGGTCAAG TGGCAGGGTC ATGAGGGAGA	60
	CATCGATAAG GGTGCTCCTT ATGCTCCCTG CTCTGGAATC CACCAGCGGG CTATCTGCGT	120
	TTATGGGGCT GGGGACTAGA ATTGGATGCT TCAAAACCAT CACCTGTTGG CCAACAAGTT	180
50	TGACCCAAAG GTAGATGATA ATGCTCTTCA GTGCTTAGAA GAATACCTAC GTTATAAGGG	240
	CCATTCTATT GGGACCTGAA CTTTGAAGAC CACAMTATG AAGAGGCGTT GCTTACCYGT	300
55	TGGGGGCCAA GAGGCATGTT ACCAAACATG GYYCARGAAM YTTGGYKGGG AMCARKKKKG	360
	GKKGGGARRM CMRGGGYTTG SCAAWTCSK KGGCMWCCYT TTAGGGTAAR RRGGGCKGTW	420
	ATTAGATGTG GGGTAAAGTA GGATCTTTTG CCCTTGCAAA TTGCTGCCTT GGGTGAATGY	480
60	TGCTTGTTCC TTCTCMACCC CTAACCTAG TAGTTCCTCC ACTAACTTTC TCACTAAGTG	540



AGAATGAGAA CTGCTGTGAT AGGGAGAGTG AAGGAGGGAT ATGTGGTAGA GCACTTGATT 600  
TCAGTTGAAT GCGTGTGGT AGCTTTTCCA TTCTGTGGAG CTGCCGTTCC TAATAATTCC 660  
AGGTTTGGTA GCGTGGAGGA GAACTTTGAT GGAAAGAGAA CCTTCCCTTC TGTACTGTTA 720  
ACTTAAAAAT AAATAGCTCC TGATTCAAAG TAAAAA AAAA AAAA 777

10

(2) INFORMATION FOR SEQ ID NO: 182:

15

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 791 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 182:

GGCACAGATA ACTATGTACA TGTATTCCTT AAATGTTTTT TTAAGTTTTA TATTCTTGGC 60  
ACTGGTCTTC AAATGTGTAC ATGTGTGCCA GGGAGCAAAT GCCTTCTTGT TTCTGAAATT 120  
GGTCTTTTAG ACTGTTCTTT TTTCCCATCT TCTCACCTCC TGCCCTCCT TCAGGGTACT 180  
TCCGTGGCCA GAACCCCTCC AGGTCAGAGG CAGAAGAGAA GCCTCATGGG TCACAGCAGC 240  
AGATGTGGGC TGGAGATCTA TTCATTGGT TTTGGCTTGA ATTTTCTGRA TGGTTTACTT 300  
GATCYTGGGA AAGANATATC TTGCCAGGAA AAATGATAGN CCTTGACAAT GTTGAATGAT 360  
CCTGCACCAC CTGAAAGAC ATTTCTAATA TGGTTTGTCA GGCAAAGTGG TTAGTAGTCA 420  
TTTGTGGCCT GAGGTAGAAG TCCTCAGAAA TCACGAGACT TCACTGATAA AATGCTGACT 480  
TGCCCTGGA CTGGGCTCTG TGAGAGTGGC CTCTGCACT GTGCACAGTA GGTGTGAACA 540  
CACCACACCT ACAGGGACCA CGTGGTGGC TGTGGACTAG CGGCCAAGCT CCCTGCAGGC 600  
CCACTAATAG AATTCAGCTT TTAGCATGGG CTGTTTCATA CTGTTCTGAT GAAACTGATT 660  
TGGTTTCTTT CCTCCATACC CCTTCTGCAT TTCAGTGT TTGTTTAGTT TTCCTGGTTT 720  
TTAATTATAA CTACAAAATA AAATCTTTAG GCTATTCACC TTAGCTTAGT AAAAAAAAAA 780  
AAAAAAACT C 791

50

55

(2) INFORMATION FOR SEQ ID NO: 183:

60

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1405 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear



(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 184:

5	GTCATGCAGT GCGCCGAGA ACTGTGCTCT TTAGGCCGA CGCTAGGGC CCGAAGGGA	60
	AACTGCGAGG CGAAGGTGAC CGGGACCGA GCATTTCAGA TCTGCTCGGT AGACCTGGTG	120
	CACCACCACC ATGTTGGCTG CAAGGCTGGT GTGTCTCCG AACTACCTT CTAGGGTTTT	180
10	CCACCCAGCT TTCACCAAGG CCTCCCTGT TGTGAAGAAT TCCATCACGA AGAATCAATG	240
	GCTGTTAACA CCTAGCAGG AATATGCCAC CAAAACAAGA ATGGGATCC GCGTGGGAG	300
15	AACTGGCCAA GAACTCAAAG AGGCAGCATT GGAACCATG ATGGAAGAAA TATTTAAAT	360
	TGATCAGATG GGAAGATGGT TTGTTGCTGG AGGGGCTGCT GTTGGTCTTG GAGCATTGTG	420
	CTACTATGGC TTGGGACTGT CTAATGAGAT TGGAGCTATT GAAAAGGCTG TAATTTGGCC	480
20	TCAGTATGTC AAGGATAGAA TTCATTCCAC CTATATGTAC TTAGCAGGGA GTATTGGTTT	540
	AACAGCTTTG TCTGCCATAG CAATCAGCAG AACGCCTGTT CTCATGAACT TCATGATGAG	600
25	AGGCTCTTGG GTGACAATTG GTGTGACCTT TGCAGCCATG GTTGGAGCTG GAATGCTGGT	660
	ACGATCAATA CCATATGACC AGAGCCCAGG CCCAAAGCAT CTGCTTGGT TGCTACATTC	720
	TGGTGTGATG GGTGCAGTGG TGGCTCCTCT GACAATAITA GGGGGTCTC TTCTCATCAG	780
30	AGCTGCATGG TACACAGCTG GCATTGTGGG AGGCCTCTCC ACTGTGGCCA TGTGTGCGCC	840
	CAGTGAAAAG TTTCTGAACA TGGGTGCACC CCTGGGAGTG GGCCTGGGTC TCGTCTTTGT	900
35	GTCTCATTG GGATCTATGT TTCTTCCACC TACCACCGTG GCTGGTGCCA CTCTTTACTC	960
	AGTGGCAATG TACGGTGGAT TAGTTCTTTT CAGCATGTTT CTCTGTATG ATACCCAGAA	1020
	AGTAATCAAG CGTGCAGAAG TATCACAAT GTATGGAGTT CAAAATATG ATCCCATTA	1080
40	CTCGATGCTG AGTATCTACA TGATACATT AAATATATTT ATGCGAGTTG CAACTATGCT	1140
	GGCAACTGGA GGCAACAGAA AGAAATGAAG TGAATCAGCT TCTGGCTTCT CTGCTACATC	1200
45	AAATATCTTG TTTAATGGG CAGATATGCA TTAAATAGTT TGTACAAGCA GCTTTCGTTG	1260
	AAGTTTAGAA GATAAGAAAC ATGTATCAT ATTTAAATGT TCCGTAATG TGATGCCTCA	1320
	GGTCTGCCTT TTTTCTGGA GAATAAATGC AGTAATCCTC TCCCAAATAA GCACACACAT	1380
50	TTTCAATTCT CATGTTTGG TGAATTTAAA ATGTTTGGT GAATGTGAAA ACTAAAGTTT	1440
	GTGTATGAG AATGTAAGTC TTTTCTTAC TTTAAATTT AGTAGTTCA CTGAGTAACT	1500
55	AAAATTTAGC AAACCTGTGT TTGCATATTT TTTKGGAGTG CAGMMTAWTG TAATTARACC	1560
	ATTCCAGTAA NAGTGTNTTT AAAGTTGNTC TATATN	1596

## (2) INFORMATION FOR SEQ ID NO: 185:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2293 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 185:

5	GCGCAGAGCC CGYACGAGCA GGACGACGAC GACAAGGGCG ACTCCAAGGA AACGCGGCTG	60
15	ACCCTGATGG AGGAAGTGCT CCTGCTGGGC CTCAAGGACC GCGARGGTTA CACATCATTT	120
	TGGAATGACT GTATATCATC TGGATTACGT GGCTGTATGT TAATTGAATT AGCATTGAGA	180
	GGAAGGTAC AACTAGAGGC TTGTGGAATG AGACGTAAAA GTCTATTAAC AAGAAAGGTA	240
20	ATCTGTAAGT CAGATGCTCC AACAGGGGAT GTTCTTCTTG ATGAAGCTCT GAAGCATGTT	300
	AAGGAACTC AGCCTCCAGA AACGGTCCAG AACTGGATTG AATTACTTAG TGGTGAGACA	360
25	TGGAATCCAT TAAAAATGCA TTATCAGTTA AGAAATGTAC GGAACGATT AGCTAAAAAC	420
	CTGGTGAAA AGGGTGTATT GACAACAGAG AAACAGAACT TCCTACTTTT TGACATGACA	480
	ACACATCCCC TCACCAATAA CAACATTAAG CAGCGCCTCA TCAAGAAAGT ACAGGAAGCC	540
30	GTCTTGACA AATGGGTGAA TGACCTCAC CGCATGGACA GCGCTTGCT GGCCCTCATT	600
	TACCTGGCTC ATGCCTCGGA CGTCTGGAG AATGCTTTTG CTCTCTTCT GGACGAGCAG	660
35	TATGATTGG CTACCAAGAG AGTGGGCGAG CTCTCGACT TAGACCTGA AGTGAATGT	720
	CTGAAGGCCA ACACCAATGA GGTCTGTGG GCGGTGGTG CCGCGTTCAC CAAGTAACTC	780
	TGCTCGGGT GAACCATCTT CCTTCTCTC AAGTAAACCA GTAGTTTTTC TTCTGTGAC	840
40	TTCTGGTTTT CTGTAATTTG TACTTCCCA CACTATAATT GGCTTCTGTT TTACAAAATG	900
	GTGGGTGGCT TTTTCTTTT TGTACGTGTA CAGGATTCTG CTGGTACGAG AGGCCTTCCT	960
45	CITTCGTFTT TAAAAAAG TTTTACTGCC ATATTGGCAT TCCATTCCCT GTTGCCATCC	1020
	TCACTGTAC CTGTTTTGGG TTTCTGGTCT ACTTTGACTT TCAAAGTACC TCCAGCCTCC	1080
	TCATACGCAC AGCTTTTGGG TGACCTCAGC TTGAGTTTCT CCATATGTGC ATGTACATCT	1140
50	AGCATCTGC CTACAGTTCA GACAGAAGTC AAAAAAGGC CTTCAACTCA CCAAAGGTAA	1200
	ATATCTGTAT CTATTAGGAC ATTTTGTACA TAGACTTCAG TTGAGATGTA TACTTAGCAA	1260
55	AATTATTTT AAATTGAAAC AGCACAGTAA ATACTTAATA TAAATGTCC CTTGGATTTT	1320
	GCTTCCCATG TAAATCTATT GTATTATTAC ACTTGTATA ATTTAACTA TAAAGGTCCA	1380
	ATTGTTTCAC AGAGCCAGT TGGGATGGG TGCAATCCAT TTATGCTGTA TATAGTTTGA	1440
60	ATTATATATA AATTACCCCT TCTCTGGCC ACCCTGCTC CCATCTTAGT ATTTTGCAAG	1500

	ATCTAATCAG TTGTACACCT GGTGCCCTC GCTTGCTTCA ATCATGGTTA TTTGATGGCA	1560
5	AAATCGACCT CTGTGCGCTG AAGGAGAGAG AAAAGATGTG TGCTGATTG GTCCTGGGAT	1620
	TTTTTGAGCT GTGCCATTTA TGGTACTCTT TGCCTATGCA TCCCCTTTTT AGATTTTTTT	1680
	TAAATTTTAT CTACTGTCTT TTATAATTTT TATTGGGAAG AGGCTTGTGA CCAGTACCAA	1740
10	TCTTGAGTTT CTTTTTCTGT CCACAAGTAA ATTAATATCT GCTCTGAAAT GTCATTTATC	1800
	TACTCACACA TTCTTGGGGA AAAAAATCAA ATGTCAGTCC TAGCAGATGT TGCATGTAAA	1860
15	TTGGTAGCAA GTAATGATTA CAACCAGAG GATTAAGAAT TTTGTAACAG AAAGCTCTAT	1920
	GTTTTAAITTT TTTATATACA ATTAGGATAA TTAGCATTGT CAGACTATAA ACCTTTGCTT	1980
	TTTAAAGTTT ATTTTACTA TTCTTTATC ACTTTATGT ATCATCACCA TTGGTTTCAT	2040
20	AATGTAAATA CTATATGTG AACAAATTAA ATGTCAAAAT TTTTATTAC CATAGTCCAT	2100
	GTTAATAGTG GGGCTTTCAG GTGTTTAGAG ATTTTTTTTG TTGTTGTTAA CATTCATTGC	2160
25	AAAAGTACTA GATGGTGTAT AACTCTAGAG TTGAATTTTA AGGGATTCCC TAATATGTAT	2220
	ACTATCTTTT TATCTGAAGT AATAAATAAA CAATGATCTT GAAAGTGCCY RAAAMAAAAA	2280
30	AAAAAAAAAA AAA	2293

## (2) INFORMATION FOR SEQ ID NO: 186:

- 35 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1212 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - 40 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 186:

	GGCACGAGGC GAGCCGGCGC ACCGTACGCT GGGACGTGTG GTTTCAGCTC GTGCGCCTCC	60
45	CCGTGGGTTT GCGACGTTTA GCGACTATTG CGCCTGCGCC ACGCCGGCTG CGAGACTGGG	120
	GCCGTGGCTG CTGGTCCCGG GTGATGCTAG GCGGCTCCCT GGGCTCCAGG CTGTTGCGGG	180
50	GTGTAGGTGG GAGTCACGGA CGGTTGCGGG CCCGAGGTGT CCGCGAAGGT GGCACACATG	240
	GGCGGCAGGG GAGAGCATGG CTCAGCGGAT GGTCTGGGTG GACCTGGAGA TGACAGGATT	300
	GGACATTGAG AAGGACCAGA TTATTGAGAT GGCCTGTCTG ATAAGTACT CTGATCTCAA	360
55	CATTTTGGCT GAAGGTCTTA ACCTGATTAT AAAACAACCA GATGAGTTGC TGGACAGCAT	420
	GTCAGATTGG TGTAAAGAGC ATCACGGGAA GTCTGGCCTT ACCAAGGCAG TGAAGGAGAG	480
60	TACAATTACA TTGCAGCAGG CAGAGTATGA ATTTCTGTCC TTTGTACGAC AGCAGACTCC	540

	TCCAGGGCTC TGTCCACTTG CAGGAAATTC AGTTCATGAA GATAAGAAGT TTCTTGACAA	600
	ATACATGCCC CAGTTCATGA AACATCTTCA TTATAGAATA ATTGATGTGA GCACGTGTAA	660
5	AGAACTGTGC AGACGCTGGT ATCCAGAAGA ATATGAATTT GCACCAAAGA AGGCTGCTTC	720
	TCATAGGGCA CTTGATGACA TTAGTGAAAG CATCAAAGAG CTTCAGTTTT ACCGAAATAA	780
10	CATCTTCAAG AAAAAAATAG ATGAAAAGAA GAGGAAAATT ATAGAAAATG GGGAAAATGA	840
	GAAGACCGTG AGTTGATGCC AGTTATCATG CTGCCACTAC ATCGTTATCT GGAGGCAACT	900
	TCTGGTGGTT TTTTCTCTC ACGCTGATGG CTTGGCAGAG CACCTTCGGT TAACCTGCAT	960
15	CTCCAGATTG ATTACTCAAG CAGACAGCAC ACGAAATACT ATTTTCTCC TAATATGCTG	1020
	TTTCCATTAT GACACAGCAG CTCCTTGTGTA AGTACCAGGT CATGTCCATC CCTTGGTACA	1080
20	TATATGCATT TGCTTTTAAA CCATTCTTTT TGYTTAAATA AATAAATAAG TAAATAAAGC	1140
	TAGTTCTATT GAAATGCAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA	1200
	AAAAAAAAAA AN	1212

25

(2) INFORMATION FOR SEQ ID NO: 187:

30

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1605 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 187:

	GCTTCCGGAA GTTGCTTTTG TCCAAACATC CGGGCTTCTC CTTTGTGTGT TCCGGCCGAT	60
40	CCCACCTCTC CTCGACCCTG GACGTCTACC TTCCGGAGGC CCACATCTTG CCCACTCCGC	120
	GCGCGGGGCT AGCGCGGGTT TCAGCGACGG GAGCCCTCAA GGGACATGGC AACTACAGCG	180
45	GCGCGGGCGG GCGCGCCCCG AAATGGAGCT GGCCCGAAT GGGGAGGGTT CGAAGAAAAC	240
	ATCCAGGGCG GAGGCTCAGC TGTGATTGAC ATGGAGAACA TGGATGATAC CTCAGGCTCT	300
	AGCTTCGAGG ATATGGGTGA GCTGCATCAG CGCTGCGCG AGGAAGAAGT AGACGCTGAT	360
50	GCAGCTGATG CAGCTGCTGC TGAAGAGGAG GATGGAGAGT TCCTGGGCAT GAAGGGCTTT	420
	AAGGGACAGC TGAGCCGGCA GGTGGCAGAT CAGATGTGGC AGGCTGGGAA AAGACAAGCC	480
55	TCCAGGGCCT TCAGCTTGTA CGCCAACATC GACATCTTCA GACCCTACTT TGATGTGGAG	540
	CCTGCTCAGG TCGCAACAGG GCTCCTGGAG TCCATGATCC CTATCAAGAT GGTCAACTTC	600
	CCCCAGAAAA TTGCAGGTGA ACTCTATGGA CCTCTCATGC TGGTCTTCAC TCTGGTTGCT	660
60	ATCCTACTCC ATGGGATGAA GACGTCTGAC ACTATTATCC GGGAGGGCAC CCTGATGGGC	720

	ACAGCCATTG GCACCTGCTT CGGCTACTGG CTGGGAGTCT CATCCTTCAT TTACTTCCTT	780
5	GCCTACCTGT GCAACGCCCA GATCACCATG CTGCAGATGT TGGCACTGCT GGGCTATGGC	840
	CTCTTTGGGC ATTGCATTGT CCTGTTCATC ACCTATAATA TCCACCTCCA CGCCCTCTTC	900
	TACCTCTTCT GGCTGTTGGT GGGTGGACTG TCCACACTGC GCATGGTAGC AGTGTGGTG	960
10	TCTCGGACCG TGGGCCCCAC ACAGCGGCTG CTCCTCTGTG GCACCTGGC TGCCCTACAC	1020
	ATGCTCTTCC TGCTCTATCT GCATTTTGCC TACCACAAAG TGGTAGAGG GATCCTGGAC	1080
15	ACACTGGAGG GCCCCAACAT CCGGCCCATC CAGAGGGTCC CCAGAGACAT CCTGCCATG	1140
	CTCCCTGCTG CTGGGCTTCC CACCACCGTC CTCAACGCCA CAGCCAAAGC TGTGCGGTG	1200
	ACCCTGCAGT CACACTGACC CCACCTGAAA TTCTTGCCA GTCTCTTTC CCGCAGCTGC	1260
20	AGAGAGGAGG AAGACTATTA AAGGACAGTC CTGATGACAT GTTTCGTAGA TGGGGTTTGC	1320
	AGCTGCCACT GAGCTGTAGC TGCCTAAGTA CCTCCTTGAT GCNIGTCGGC ACTTCTGAAA	1380
25	GGCACAAGGC CAAGAACTCC TGGCCAGGAC TGCAAGGCTC TGCAGCCAAT GCAGAAAATG	1440
	GGTCAGCTCC TTTGAGAACC CCTCCCCACC TACCCCTTCC TTCTCTTTA TCTCTCCAC	1500
	ATTGTCTTGC TAAATATAGA CTTGTAATT AAAATGTTGA TTGAAGTCTG GAAAAAATA	1560
30	AAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAC TCGAG	1605

35 (2) INFORMATION FOR SEQ ID NO: 188:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1516 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 188:

45	ATTCGGCATG AGGGGGTCAC GTGGTGGCTG GGCCGGGGAA ATGGCGGCTT CAGGAGAGAG	60
	CGGCACTTCA GCGCGCGGAG GCAGCACCGA GGAAGCATTT ATGACCTTCT ACAGTGAGGT	120
50	GAAACAAATA GAGAAGAGAG ACTCGTTTCT AACTTCGAAA AATCAGATTG AAAGACTGAC	180
	CCGTCTGGT TCCTCTTACT TCAATTTGAA CCCATTTGAG GTTCTTCAGA TAGATCCTGA	240
	AGTTACAGAT GAAGAAATAA AAAAGAGGTT TCGGCAGTTA TCCATCTTGG TGCATCCTGA	300
55	CAAAAATCAA GATGATGCTG ACAGAGCACA AAAGGCTTTT GAAGCTGTGG ACAAAGCTTA	360
	CAAGTTGCTA CTGGATCAGG AGCAAAAGAA GAGGGCCCTG GATGTAATTC AGGCAGGAAA	420
60	AGAATACGTG GAACACACTG TGAAAGAGCG AAAAAACAA TTAAAGAAGG AAGGAAAACC	480

	TACAATTGTA GAGGAGGATG ATCCTGAGCT GTTCAAACAA GCTGTATATA AACAGACAAT	540
	GAAACTCTTT GCAGAGCTGG AAATTAAAAG GAAAGAGAGA GAAGCCAAAG AGATGCATGA	600
5	AAGGAAACGA CAAAGGGAAG AAGAGATTGA AGCTCAAGAA AAAGCCAAAC GGGAAAGAGA	660
	GTGGCAGAAA AACTTTGAGG AAAGTCGAGA TGGTCGTGTG GACAGCTGGC GAAACTTCCA	720
10	AGCCAATACG AAGGGAAGA AAGAGAAGAA AAATCGGACC TTCCTGAGAC CACCGAAAGT	780
	AAAAATGGAG CAACGTGAGT GACCGCCCAA GGTACAGGC ACAGAACCTT TCCCCTGCTA	840
	TCTCCCTTCC TGCTTCGAAG GACTCATTCT TTCTCCAC TTCCACCCCA ACATAGAGTA	900
15	GTATTTGCTT TTTAGTCCAT TTTGTTTCA ATACGATTTA ATATCGATCA GAGTAATTCT	960
	TTTGATACATT GAAATGAGG GCTTGGTTTA AAAAAAGACC TTTCCCTCTC CCGCCCTTA	1020
20	GAACAACCAG TATTAGAAGG TGCCACCATT GGTGCTGCTT TCTCTTCCA CAGCCTGTAA	1080
	CTCAGTGT TTGTACTTCA TGAATTGTGA TGGTTAGAAA CTTCGTGGAT AGTTTGTGGA	1140
	AATCATCCAA TTAACATAC TGCTTAAAC AGTGTGCTG TGACTTCAGA GACAAGCCTG	1200
25	GAAGGGCAC CTAGGAAGC CCCTTCGCTT CAGTTGCTCG CTCTGGGTG TGCTCCCTTC	1260
	GAAGGCCAG ATAAGACAGG GAACACTTGT GAGCACACAG AGCAGCATCT GATGCCCTGT	1320
30	GGTGTGTC ATGTGCCCC TGCTACTGA CCAATCAGTG TGGCATGAGG CCCACGCCAC	1380
	CCAAACCTTT CACTTTCCAA AGAGCTAGCC GTCCTCCACC CAGTACCATG TCCTAGCCTG	1440
	TCTGCATTG TTAGTGGTAA TATCTTTAT GTATAATAAA TTTTATACC CAAAAAAAAA	1500
35	AAAAAAAAA ACTCGA	1516

40 (2) INFORMATION FOR SEQ ID NO: 189:

(i) SEQUENCE CHARACTERISTICS:

- 45 (A) LENGTH: 681 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 189:

50	GCTCCCATGT TGCTGGCTGT CCGTACATCA CCCTGTCCCC TGCAGGAGGG GGCTACAGGC	60
	CATCTCCCTC CTGTAGGCCT CTGACTCCCC TCCACTTTTG GGCCCTCAGC TTATCTCGGG	120
55	CAGGGGACCA TTGCAGCATC CTCCCCCTCT CNGGACTCAA GGTGCTGAGG TATAAGCCCT	180
	GGGCCCCAGA TCCCTGRTKA CACCTTCCTG GAGAAGACTC TCAAAGTGA CTGTATATTT	240
	GAGTTCACCA GCAATAACTC CCCACACTCG AAGCAGGTCC AAACCCMAGG ATCCCAGGCT	300
60	CCTTGGGCTC TGTGGCACTG TCTTCCCAAG ATCCTTCCTG TTGCACAATG GGAAACCTAA	360



5 GAGGAAAAAG ACAGGGGCCT GCTTGCCAG CCATGCGAGG GATTCCATGC CCACCTGCCC 420  
 TCTGYCTGCC TCGCTGGAAT GTGGGCCCT GCTCCCGTC AGGTGTGCT GTCTCTGACC 480  
 TATGTTTACA TCCCCGAGG GTTCTGCCT CCTCCCCACC CAGGTCAGG TGTGGTCCAG 540  
 CAGCTTGCTG TGGGGTGCTG ACATGTGTCA CCACTGCCCC CCTTGCCCC GGGGGGTCA 600  
 10 TGGTCTCTC CTGGATGCTG CTCCTTGAAT YTTTTTYYT GAWAAACCYT TTAMAATTAA 660  
 AAAAAAAAAA AAAAACTCG A 681

15

(2) INFORMATION FOR SEQ ID NO: 190:

20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1014 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 190:

GCCTCAAGCC ACGCATATGA TAATTTCTG GAACATTCAA ATTCACTGTT TCTACAGCCA 60  
 30 GTTAGTCTAC AAACCATTC AGCAGACCA TCAAACCAGA GTCTGCCACT TTTTGTATC 120  
 GCTGGATGAT TGCTGGGCAA AGGTGGCCTT TTAGAGCTCT TAAAGCCCA CAAAAGGCT 180  
 ATTCGTAGAG CCACAGTCAA CACATTGGT TATATTGCAA AGGCCATTGG CCTCATGATG 240  
 35 TATTGGCTAC ACTTCTGAAC AACCTCAAAG TTCAAGAAAG GCAGAACAGA GTTTGTACCA 300  
 CTGTAGCAAT AGCTATTGTT GCAGAAACAT GTTCACCTT TACAGTACTC CCTGCCTTAA 360  
 TGAATGAATA CAGAGTTCTT GAACTGAATG TTCAAATGG AGTGTTAAA TCGCTTCTCT 420  
 40 TCTTGTTTGA ATATATTGGT GAAATGGGA AAGACTACAT TTATGCCGTA ACACCGTTAC 480  
 TTGAAGATGC TTTAATGGAT AGAGACCTTG TACACAGACA GACGGCTAGT GCAGTGGTAC 540  
 45 AGCACATGTC ACTTGGGTT TATGGATTG GTTGTGAAGA TTCGCTGAAT CACTTGTGA 600  
 ACTATGTATG GCCCAATGTR TTTGAGACAT CTCCTCATGT AATTCAGGCA GTTATGGGAG 660  
 CCTAGAGGG CTGAGAGTT GCTATTGGAC CATGTAGAAT GTTGCAATAT TGTTTACAGG 720  
 50 GTCTGTTTCA CCCAGCCCGG AAAGTCAGAG ATGTATATTG GAAAATTAC AACTCCATCT 780  
 ACATTGGTTC CCAGGACGCT CTCATAGCAC ATTACCAAG AATCTACCA CGATGATAAG 840  
 55 RACACCTATA TTCGTTATGA ACTTGACTAT ATCTTATAAT TTTATTGTTW ATTTKGTGKT 900  
 TAATGCACAS TACTTCACAC CTTAACTTG CTTTGATTG GTGATGTAAA CTTTTAAACA 960  
 60 TTGCAGATCA GTGTAGGACT GGTCCATAGG GGAAGAGCTA GGAANTCCAT AGGC 1014

(2) INFORMATION FOR SEQ ID NO: 191:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2779 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 191:

15	TCGCAGCAGG GTGTGTCCAG ATGGTCAGTC TCTGGTGGCT AGCCTGTCTCT GACAGGGGAG	60
	AGTTAAGCTC CCGYTCTCCA CCGTGCCGGC TGGCCAGGTG GGCTGAGGGT GACCGAGAGA	120
	CCAGAACCTG CTTGCTGGAG CTTAGTGCTC AGAGCTGGGG AGGGAGGTTT CGCCGCTCCT	180
20	CTGCTGTCTAG CGCCGGCAGC CCCTCCCGGC TTCACTTCCT CCCGCAGCCC CTGCTACTGA	240
	GAAGCTCCGG GATCCCAGCA GCCGCCACGC CCTGGCCTCA GCCTGCGGGG CTCCAGTCAG	300
25	GCCAACACCG ACGCGCANTG GGAGGAAGAC AGGACCCTTG ACATCTCCAT CTGCACAGAG	360
	GTCTTGCTG GACCGAGCAG CCTCTCTCTC CTAGGATGAC CTCACCTCC AGCTCTCCAG	420
	TTTTCAGGTT GGAGACATTA GATGGAGGCC AAGAAGATGG CTCTGAGGCG GACAGAGGAA	480
30	AGCTGGATTT TGGGAGCGGG CTGCCTCCCA TGGAGTCACA GTTCCAGGGC GAGGACCGGA	540
	AATTGCGCCC TCAGATAAGA GTCAACCTCA ACTACCGAAA GGAACAGGT GCCAGTCAGC	600
35	CGGATCCAAA CCGATTTGAC CGAGATCGGC TCTTCAATGC GGTCTCCCGG GGTGTCCCCG	660
	AGGATCTGGC TGGACTTCCA GAGTACCTGA GCAAGACCAG CAAGTACCTC ACCGACTCGG	720
	AATACACAGA GGGCTCCACA GGTAAAGCGT GCCTGATGAA GGCTGTGCTG AACCTTAAGG	780
40	ACGGGGTCAA TGCTTGCAAT CTGCCACTGC TGCAGATCGA CCGGGACTCT GGCAATCCTC	840
	AGCCCCCTGGT AAATGCCCAG TGCACAGATG ACTATTACCG AGGCCACAGC GCTCTGCACA	900
45	TGCCCATGTA GAAGAGGAGW CTGCAGTGTG TGAAGCTCCT GGTGGAGAAT GGGGCCAATG	960
	TGCATGCCCG GGTCTGCGGC GCTTCTTCCA GAAGGGCCAA GGGACTTGCT TTTATTTCCG	1020
	TGAGCTACCC CTCTYTTTGG CCGCTTGCAC CAAGCAGTGG GATGTGGTAA GCTACCTCCT	1080
50	GGAGAACCCA CACCAGCCCG CCAGCCTGCA GGCCTGACT CCCAGGGCAA CACAGTCTGT	1140
	CATGCCCTAG TGATGATCTC GGACAACTCA GCTGAGAACA TTGCACTGGT GACCAGCATG	1200
55	TATGATGGGC TCCTCCAAGC TGGGGCCCGC CTCTGCCCTA CCGTGCACTG TGAGGACATC	1260
	CGCAACCTGC AGGATCTCAC GCCTCTGAAG CTGGCCGCCA AGGAGGGCAA GATCGAGATT	1320
	TTCAGGCACA TCCTGCAGCG GGAGTTTTC A GACTGAGCC ACCTTTCCCG AAAGTTTACC	1380
60	GAGTGGTGCT ATGGGCTGTG CCGGTGTGCG CTGTATGACC TGGCTTCTGT GGACAGCTGT	1440

	GAGGAGAACT CAGTGCTGGA GATCATTGCC TTTCATTGCA AGAGCCCGCA CCGACACCGA	1500
5	ATGGTCGTTT TGGAGCCCCCT GAACAACTG CTGCAGGCGA AATGGGATCT GCTCATCCCC	1560
	AAGTCTTCT TAAACTTCCT GTGTAATCTG ATCTACATGT TCATCTTCAC CGCTGTTGCC	1620
	TACCATCAGC CTACCTGAA GAAGCAGGCC GCCCTCACC TGAAAGCGGA GGTGGAAC	1680
10	TCCATGCTGC TGACGGGCA CATCCTTATC CTGCTAGGGG GGATCTACCT CCTCGTGGGC	1740
	CAGCTGTGGT ACTTCTGGCG GCGCCAGTG TTCACTGGA TCTCGTTCAT AGACAGCTAC	1800
15	TTTGAAATCC TCTTCTGTT CCARGCCCTG CTCACAGTGG TGTCCARGT GCTGTGTTTC	1860
	CTGGSCATCG AGTGGTACCT GCCCTGCTT GTGTCTGGCG TGGTCTGGG CTGGCTGAAC	1920
	CTGCTTACT ATACACGTGG CTTCCAGCAC ACAGGCATCT ACAGTGTCTAT GATCCAGAAG	1980
20	CCCTGGTGAG CCTGAGCCAG GANNTTGGCG CCCCGAAGCT CCTACAGGCC CCAATGCCAC	2040
	AGAGTCAGTG CAGCCATGG AGGGACAGGA KGACGARGGC AACGGGGCCC AGTACAGGGG	2100
25	TATCCTGGAA GCCTCCTGG AGCTCTTCAA ATTCAACATC GGCATGGGCG AGCTGGCCTT	2160
	CCAGGARCAG CTGCACTTCC GCGGCATGGT GCTGCTGCTG CTGCTGGSCT ACGTGTGCT	2220
	CACCTACATC CTGCTGCTCA ACATGCTCAT CGCCCTCATG AGCGAGACCG TCAACAGTGT	2280
30	CGCCACTGAC AGCTGGAGCA TCTGGAAGCT GCAGAAAGCC ATCTCTGTCC TGGAGATGGA	2340
	GAATGGCTAT TGGTGGTGCA GGAAGAAGCA GCGGGCAGGT GTGATGCTGA CCGTTGGCAC	2400
35	TAAGCCAGAT GGCAGCCCSG ATGAGCGCTG GTGCTTCAGG GTGGAGGAGG TGAAGTGGGC	2460
	TTTATGGGAG CAGACGCTGC CTACGCTGIG TGAGGACCCG TCAGGGGCAG GTGTCCCTCG	2520
	AACTCTCGAG AACCTGTCC TGGCTTCCCC TCCAAGGAG GATGAGGATG GTGCCCTCTGA	2580
40	GGAAACTAT GTGCCCGTCC AGCTCCTCCA GTCCAACTGA TGGCCAGAT GCAGCAGGAG	2640
	GCCAGAGGAC AGAGCAGAGG ATCTTTCCAA CCACATCTGC TGGCTCTGGG GTCCCAGTGA	2700
45	ATTCTGGTGG CAAATATATA TTTTCACTAA CTCAAAAAA AAAAAAAAAA AAAAAAAAAA	2760
	AAAAAAAAA AAAAAAGGC	2779

50

(2) INFORMATION FOR SEQ ID NO: 192:

(i) SEQUENCE CHARACTERISTICS:

55

(A) LENGTH: 1923 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 192:

	ACCCGCTCCG CTCGCTCCG CTCGGCCCCG CGCGCCCCGT CAACATGATC CGCTGCGGCC	60
	TGGCCTGCCA GCGCTGCCG TGGATCCTGC CCTGCTCCT ACTCAGCGCC ATCGCCTTCG	120
5	ACATCATCGC GCTGGCCGGC CGCGGCTGGT TGCAGTCTAG CGACCACGGC CAGACGTCCT	180
	CGCTGTGGTG GAAATGCTCC CAAGAGGGCG GCGGCAGCGG GTCCTACGAG GAGGGCTGTC	240
10	AGAGCCTCAT GGAGTACGG TGGGTAGAG CAGCGGCTGC CATGCTCTTC TGTGGCTTCA	300
	TCATCCTGGT GATCTGTTTC ATCCTCTCCT TCTTCGCCCT CTGTGGACCC CAGATGCTTG	360
	TCTTCTGAG AGTGATTGGA GGTCTCCTMG CCTTGGCTGC TGTGTCCAG ATCATCTCCC	420
15	TGGTAATTTA CCCCCTGAAG TACACCCAGA CCTTCACCTT TCATGCCAAC CSTGCTGTCA	480
	CTTACATCTA TAACTGGGCC TACGGCTTTG GGTGGGCAGC CACGATTATC CTGATYGGCT	540
20	GTGCTTCTT CTTCTGCTGC CTCGCCAACT ACGAAGATGA CCTTCTGGGC AATGCCAAGC	600
	CCAGGTACTT CTACACATCT GCCTAACTTG GGAATGAATG TGGGAGAAAA TCGCTGCTGC	660
	TGAGATGGAC TCCAGAAGAA GAACTGTTT CTCCAGGCGA CTTTGAACCC ATTTTTTGGC	720
25	AGTGTTCATA TTATTAACT AGTCAAAAAT GCTAAAATAA TTTGGGAGAA AATATTTTTT	780
	AAGTAGTGT ATAGTTTCAT GTTTATCTTT TATTATGTTT TGTGAAGTTG TGTCTTTTCA	840
30	CTAATTACCT ATACTATGCC AATATTTCCT TATATCTATC CATAACATTT ATACTACATT	900
	TGTAAGAGAA TATGCACGTG AAACCTAACA CTTTATAAGG TAAAAATGAG GTTTCCAAGA	960
	TTTAATAATC TGATCAAGTT CTTGTATTCT CCAAATAGAA TGGACTCGGT CTGTTAAGGG	1020
35	CTAAGGAGAA GAGGAAGATA AGGTTAAAAG TTGTTAATGA CCAACATTC TAAAAGAAAT	1080
	GCAAAAAAAA AGTTATTITT CAAGCCTTCG AACTATTITTA GGAAAGCAAA ATCATTTCTCT	1140
40	AAATGCATAT CATTTGTGAG AATTTCTCAT TAATATCTTG AATCATTCAT TTCAGCTAAG	1200
	GCTTCATGTT GACTCGATAT GTCATCTAGG AAAGTACTAT TTCATGGTCC AAACCTGTTG	1260
	CCATAGTTGG TAAGGCTTTC CTTTAAGTGT GAAATATTTA GATGAAATTT TCTCTTTTAA	1320
45	AGTTCTTTAT AGGGTTAGGG TGTGGGAAAA TGCTATATTA ATAAATCTGT AGTGTPTTGT	1380
	GTTTATATGT TCAGAACCAG AGTAGACTGG ATTGAAAGAT GGACTGGGTC TAATTTATCA	1440
50	TGACTGATAG ATCTGGTTAA GTTGTGTAGT AAAGCATTAG GAGGGTCATT CTTGTCACAA	1500
	AAGTGCCACT AAAACAGCCT CAGGAGAATA AATGACTTGC TTTTCTAAAT CTCAGGTTTA	1560
	TCTGGGCTCT ATCATATAGA CAGGCTTCTG ATAGTTTGCA ACTGTAAGCA GAAACCTACA	1620
55	TATAGTTAAA ATCTGGTCT TTCTTGGTAA ACAGATTTTA AATGTCTGAT ATAAAACATG	1680
	CCACAGGAGA ATTCCGGGAT TTGAGTTTCT CTGAATAGCA TATATATGAT GCATCGGATA	1740
60	GGTCATTATG ATTTTTTACC ATTTGACTT ACATAATGAA AACCAATTCA TTTTAAATAT	1800

	CAGATTATTA TTTTGTAAGT TGTGGAAAAA GCTAATTGTA GTTTTCATTA TGAAGTTTTC	1860
	CCAATAAACC AGGTATTCTA AAAAAAAAAA AAAAAAACTN GAGGGGGGGC CCGGTACCCA	1920
5	ATT	1923
10	(2) INFORMATION FOR SEQ ID NO: 193:	
	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 2346 base pairs	
	(B) TYPE: nucleic acid	
15	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 193:	
20	AGGCTCAGGG GGACACTCTC AAAATTACAC AGCTTTTAAC AGGTGGCAGA ATGGGGGTTC	60
	AGACCCAGAT CTGGGTTCAA GTCACATCATG GTGTGATTGC GGCATTTCCTT CCCGCATCTG	120
25	GGCCTTGCCA TCTCTCTCTC CGAGTGGACA TGGAGAGGAC GGGGGCCCAG CAGCTGGATG	180
	GCTGCAGGGG ATCAAGTCTT CTCTGGGGCT GGGCACGTAN AAGAGCATGT GGCTGGTGGG	240
	CGGCATGCCT GGCTCCTCAC CTGGCAGTCT GCCTGCCCTG CTAACCGGCT GTCTCTGTGT	300
30	CCCCTAGTGC CCTCGGCTAG CATGACCCGC CTGATGCGWT SCCGCACAGC CTCTGGTTCC	360
	AGCGTCATTG TCTGGATGGC ACCCGCAGCC GCTCCCACAC CAGCGAGGGC ACCCGAAGCC	420
35	GCTCCCACAC CAGCGAGGGC ACCCGCAGCC GCTCGCACAC CAGCGAGGGG GCCCACCTGG	480
	ACATCACCCC CAACTCGGGT GCTGCTGGGA ACAGNGCCGG GCCCAAGTCC ATGGAGGTCT	540
	CCTGCTAGGC GGCTTGCCCA GCTGCCGCCC CCGGACTCTG ATCTCTGTAG TGGCCCCCTC	600
40	CTCCCCGGCC CCTTTTCGCC CCTGCGCTGC CATACTGCGC CTAAGTCCGT ATTAATCCAA	660
	AGCTTATTTT GTAAGAGTGA GCTCTGGTGG AGACAAATGA GGTCTATTAC GTGGGTGCCC	720
45	TCTCCAAAGG CGGGGTGGCG GTGGACCAA GGAAGGAAGC AAGCATCTCC GCATCGCATC	780
	CTCTTCCATT AACCAAGTGGC CGGTGCCAC TCTCTCCCC TCCCTCAGAG ACACCAAAT	840
	GCCAAAAACA AGACGCGTAC AGCACACACT TCACAAAGCC AAGCCTAGGC CGCCCTGAGC	900
50	ATCCTGGTTC AAACGGGTGC CTGGTCAGAA GGCCAGCCGC CCACTTCCCG TTTCCTCTTT	960
	AACTGAGGAG AAGCTGATCC AGTTTCCGGA AACAAATCC TTTTCTCATT TGGGGAGGGG	1020
55	GGTAATAGTG ACATGCAGGC ACCTCTTTTA AACAGGCAA ACAGGAAGGG GGAAAAGGTG	1080
	GGATTTCATG CGAGGCTAGA GGCATTTGGA ACAACAAATC TACGTAGTTA ACTTGAAGAA	1140
	ACCGATTTT AAAGTTGGTG CATCTAGAAA GCTTTGAATG CAGAAGCAA CAAGCTTGAT	1200
60	TTTCTAGCA TCCTCTTAAT GTGCAGCAA AGCAGGCRAC AAAATCTCCT GGCTTTACAG	1260

	ACAAAAATAT TTCAGCAAAC GTTGGGCATC ATGGTTTTTG AAGGCTTTAG TTCTGCTTTC	1320
5	TGCCTCTCCT CCACAGCCCC AACCTCCAC CCCTGATACA TGAGCCAGTG ATTATCTCTG	1380
	TTCAGGGAGA AGATCATTTA GATTTGTTTT GCATTCTCTA GAATGGAGGG CAACATTCCA	1440
	CAGCTGCCCT GGCTGTGATG AGTGTCTTGG CAGGGGCCGG AGTAGGAGCA CTGGGGTGGG	1500
10	GGCGGAATTG GGGTACTCG ATGTAAGGGA TTCCTTGTGTG TTGTGTGAG ATCCAGTGCA	1560
	GTGTGATTT CTGTGGATCC CAGCTTGGTT CCAGGAATTT TGTGTGATTG GCTTAAATCC	1620
15	AGTTTCAAT CTTCGACAGC TGGGCTGGAA CGTGAAGTCA GTAGCTGAAC CTGTCTGACC	1680
	CGTCAAGTT CTGTGATCCT CAGAACTCTT TGCTCTTGTG GGGGTGGGG TGGAACTCA	1740
	CGTGGGAGC GGTGGCTGAG AAAATGTAAG GATTCGTGAA TACATATTCC ATGGGACTTT	1800
20	CCTTCCCTCT CCTGCTTCT CTTTCTCTGC TCCCTAACCT TTCGCCGAAT GGGGCAGCAC	1860
	CACTGACGTT TCTGGCGGC CAGTGGGCT GCCAGGTTCC TGTACTACTG CCTGTACTT	1920
25	TTCATTTGG CTCACCGTGG ATTTTCTCAT AGGAAGTTG GTCAGAGTGA ATTGAATATT	1980
	GTAAGTCAGC CACTGGGACC CGAGGATTC TGGGACCCCG CAGTTGGGAG GAGGAAGTAG	2040
	TCCAGCCTTC CAGGTGGCGT GAGAGGCAAT GACTCGTTAC CTGCCGCCA TCACCTTGA	2100
30	GGCTTCCCT GGCCTTGAGT AGAAAAGTCG GGGATCGGG CAAGAGAGGC TGAGTACGGA	2160
	TGGAAACTA TTGTGCACAA GTCTTCCAG AGGAGTTTCT TAATGAGATA TTTGTATTTA	2220
35	TTTCCAGACC AATAAATTG TAACTTTGCA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA	2280
	AAAAAAAAA AAAAAAACT CGAGGGGGC CCGTACCCAA TTCGCCGTAT ATGATCGTAA	2340
	ACAATC	2346

40

## (2) INFORMATION FOR SEQ ID NO: 194:

45

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3054 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

50

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 194:

	TATCTGAACC ACCCTTTATT CTACATATGA TAGGCAGCAC TGAAATATCC TAACCCCTTA	60
55	AGCTCMAGGT GCCCTGTGN ACGAGCAACT GGACTATAGC AGGGCTGGGC TCTGTCTTCC	120
	TGGTCATAGG CTCACTCTTT CCCCCAAATC TTCCTCTGGA GCTTTGCAGC CAAGGTGCTA	180
60	AAAGGAATAG GTAGGAGACC TCTTCTATCT AATCCTTAAA AGCATAATGT TGAACATTCA	240

	TTCAACAGCT GATGCCCTAT AACCCCTGCC TGGATTCTT CCTATTAGGC TATAAGAAGT	300
	AGCAAGATCT TTACATAATT CAGAGTGGTT TCATTGCCTT CCTACCCCTCT CTAATGGCCC	360
5	CTCCATTAT TTGACTAAAG CATCACACAG TGGCACTAGC ATTATACCAA GAGTATGAGA	420
	AATACAGTGC TTTATGGCTC TAACATTACT GCCTTCAGTA TCAAGGCTGC CTGGAGAAAG	480
10	GATGGCAGCC TCAGGGCTTC CTTATGTCTT CCACCACAAG AGCTCCTTGA TGAAGGTCAT	540
	CTTTTCCCC TATCCTGTTC TTCCCTCCCG CGCTCCTAAT GGTACGTGGG TACCCAGGCT	600
	GGTCTTGGG CTAGGTAGTG GGGACCAAGT TCATTACCTC CCTATCAGTT CTAGCATAGT	660
15	AAACTACGGT ACCAGTGTTA GTGGGAAGAG CTGGGTTTTC CTAGTATACC CACTGCATCC	720
	TACTCCTACC TGGTCAACCC GCTGCTTCCA GGTATGGGAC CTGCTAAGTG TGGAATTACC	780
20	TGATAAGGGA GAGGGAAATA CAAGGAGGGC CTCTGGTGT TCTGGCCTCA GCCAGCTGCC	840
	CACAAGCCAT AAACCAATAA AACAAGAATA CTGAGTCAGT TTTTATCTG GGTCTCTTC	900
	ATTCCTACTG CACTTGGTGC TGCTTTGGCT GACTGGGAAC ACCCCATAAC TACAGAGTCT	960
25	GACAGGAAGA CTGGAGACTG*TCCTTCTA GCTCGGAAGT TACTGTGTAA ATAACTTTC	1020
	AGAAGTCTA CCATGAAGTG AAAATGCCAC ATTTTGCTTT ATAATTTCTA CCCATGTTGG	1080
30	GAAAACTGG CTTTTTCCCA GCCCTTTCCA GGCATAAAA CTCAACCCCT TCGATAGCAA	1140
	GTCCCATCAG CCTATTATTT TTTTAAAGAA AACTTGCACT TGTTTTCTT TTTACAGTTA	1200
	CTTCTTCTT GCCCAAAAT TATAAATCT AAGTGTAATA AAAAGTCTTA ACAACAGCTT	1260
35	CTTGCTTGTA AAAATATGTA TTATACATCT GTATTTTAA ATCTGCTCC TGAAAAATGA	1320
	CTGTCCCAT CTCCACTCAC TGCAATTGGG GCCTTTCCCA TTGGTCTGCA TGTCTTTAT	1380
40	CATTGCAGGC CAGTGGACAG AGGGAGAAGG GAGAACAGGG GTCGCCAACA CTGTGTGTC	1440
	TTTCTGACTG ATCCTGAACA AGAAAGAGTA AACTGAGGC GCTGCTCCC ATGCACAACT	1500
	CTCCAAAACA CTTATCTCTC TGCAAGAGTG GGCTTTCCAG GGTCTTACT GGAAGCAGT	1560
45	TAAGCCCCCT CCTCACCCCT TCCTTTTTC TTTCTTACT CCTTGGCTT CAAAGGATTT	1620
	TGGAAAAGAA ACAATATGCT TTACTCAT TTTCAATTC TAAATTTGCA GGGGATACTG	1680
50	AAAAATACGG CAGGTGGCCT AAGGCTGCTG TAAAGTTGAG GGGAGAGGAA ATCTTAAGAT	1740
	TACAAGATAA AAAACGAAT CCCTAAACAA AAAGAACAAT AGAACTGGTC TTCCATTTTG	1800
	CCACCTTTC TGTTCATGAC AGCTACTAAC CTGGAGACAG TAACATTTCA TTAACCAAAG	1860
55	AAAGTGGGTC ACCTGACCTC TGAAGAGCTG AGTACTCAGG CCACTCCAAT CACCCTACAA	1920
	GATGCCAAGG AGGTCCCAGG AAGTCCAGCT CCTTAAACTG ACGCTAGNCA ATAAACCTGG	1980
60	GCAAGTGAGG CAAGAGAAAT GAGGAAGAAT CCATCTGTGA GGTGACAGGC AAGGATGAAA	2040

	GACAAAGAAG GAAAAGAGTA TCAAAGGCAG AAAGGAGATC ATTTAGTTGG GTCTGAAAGG	2100
	AAAAGTCTTT GCTATCCGAC ATGTACTGCT AGTACCTGTA AGCATTITTAG GTCCCAGAAT	2160
5	GGAAAAAATA ATCAGCTATT GGTAATATAA TAATGTCCTT TCCCTGGAGT CAGTTTITTT	2220
	AAAAAGTTAA CTCTTAGTTT TTACTTGTTT AATTCTAAAA GAGAAGGGAG CTGAGGCCAT	2280
10	TCCCTGTAGG AGTAAAGATA AAAGGATAGG AAAAGATTCA AAGCTCTAAT AGAGTCACAG	2340
	CTTTCCAGG TATAAACCTT AAAATTAAGA AGTACAATAA GCAGAGGTGG AAAATGATCT	2400
	AGTTCTGAT AGCTACCCAC AGAGCAAGTG ATTTATAAAT TTGAAATCCA AACTACTTTC	2460
15	TTAATATCAC TTGGTCTCC ATTTTCCCA GGACAGGAAA TATGTCCCCC CCTAACTTTC	2520
	TTGCTTCAAA AATTAAATC CAGCATCCCA AGATCATTCT ACAAGTAATT TTGCACAGAC	2580
20	ATCTCTCAC CCCAGTGCCT GTCTGGAGCT CACCCAAGGT CACCAACAA CTGGTTGTG	2640
	AACCNACTG CCTTAACCTT CTGGGGGAGG GGGATTAGCT AGACTAGGAG ACCAGAAGTG	2700
	AATGGGAAAG GGTGAGGACT TCACAATGTT GGCCTGTCAG AGCTTGATTA GAAGCCAAGA	2760
25	CAGTGGCAGC AAAGGAAGAC TTGGCCCAGG AAAAACCTGT GGGTTGTGCT AATTCTGTC	2820
	CAGAAAATAG GGTGGACAGA AGCTTGTTGG GTGCATGGAG GAATTGGAC CTGGTTATGT	2880
30	TGTTATCTC GACTGTGAA TTTTGGTGAT GTAAAACAGA ATATTCTGTA AACCTAATGT	2940
	CTGTATAAAT AATGAGCGTT AACACAGTAA AATATTCAAT AAGAAGTCAA AAAAAAAAAA	3000
35	AAAAAATCG AGGGGGGGCC CGGTACCCAA TTTCNCAAAT AGAGATNGTA TTAC	3054

## (2) INFORMATION FOR SEQ ID NO: 195:

- 40 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 907 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

- 45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 195:

	GGCAGAGCTC GTGCGCGNAA CTTTTCCTGC TCCTGGCTGC CACCTACTGG CTGGCCGCGG	60
50	CCCTGGCCTG GGCTGCACC AGCCTGCGNG CGGCTCCCA CAGCAGCCCC CTTCGAAGCA	120
	GGTCCCCAC ACCGCGCACC TTCTGCGGGA ACGTGCTGCG CGTGCCGGGG ACCATATGGA	180
55	CGGAAGGCTT TGTGCTCACC TACAAGCTGG GTGAGCAGGG TGCCAGCAGC CTGTTGATCC	240
	TCTTGCTCC TGCTGGAGCA CGAGCGCGT TTCTGCTCCC GAGTTGGGAC TGTGGAATGG	300
	TGTGGGTGCT GTGGTCTGCT CCATCGCTGG CTCTCCCTG GGTGGGACCT TGCTGGCCAA	360
60	GCACTGGAAA CTGCTGCCTC TGTGAGGTCG GTGCTGCGCT TCCGCTCGG GGGCCTAGCC	420



5 TGTCAGACTG CCTTGGTCTT CCACCTTGGA CACCCTGGGG GCCAGCATGG ACGCTGGCAC 480  
 AATCTTGAGA GGGTCAGCCT TGCTGAGCCT ATGTCTGCAG CACTTCTTGG GARGCCTGGT 540  
 CACCACAGTC ACCTTCACTG GGAATGATGC GCTGCAGCCA GCTGGCCCCC AGGGCCTTGC 600  
 AGGCCACACA CTACAGCCTT CTGGCCACGC TGGAGCTGCT GGGGAAGCTG CTGCTGGGCA 660  
 10 CTTYTGSCGG AGGGCCTGGC TGATGGGTTG GGGCCACATC CCTGCTTCTT GCTCCTGCTC 720  
 ATCCTCTCTG CCTTCCCGT TCTGTACCTG GACCTAGCAC CCAGCACCTT TCTCTGAGCT 780  
 GAGTGGCTGG AGTGGTCAAT AAAGCCACAT GTGCCTGTGG CCCAAAAAA AAAAAAAAAA 840  
 15 AAAAAAAAAA AAAAAAATG GAGGGGGGC CCGTACCCA AATCGCCGA TATGATCGTA 900  
 AACAATC 907

20

(2) INFORMATION FOR SEQ ID NO: 196:

25

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1290 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 196:

GGCACGAGGA GGGACAGGA GTGGCAAGG GGAAGAAGCA GCTTATTGA CTAACCAGCC 60  
 35 CCTCTGTGGT CCACCAGCGT CTTGGCTTGG TGGGAGGGCT CTCAATCAGC AGGGCCCCAG 120  
 KAGGGCAAGA AGAAGTGGG CAAAGCCTGG CGCTCGGCCG CGGTGCGGC AGCTTTGCM 180  
 TCTGGAGCCA CGCCTCTCC AGGCCATGCT CCTTGAACCTT GGAAATGTCA ACCGGAGCCC 240  
 40 TTAACACCAG CCTCCAGCA TCTAATAGAC TTGAATCTAC TCTAAACGAA TATTTAATCC 300  
 AACCTCAACT ACATTGTAGC TCAGTCCAAC GACTAACCCT GAAATGGGG TGTTCCAGCC 360  
 45 TTCAGCGAGA TGGCCAAGCG GTCCCTGGG GGCTGTGGCA GCGGGCTTAT CCTTCTCTGT 420  
 TGCCAACCTT GCCGTCCGAC CTCCTCCGCC CCCATGCGGT GACCCCGTCC GTGTCTGTGT 480  
 CTGTCCATAC GTGTGAGTCC AGCTAAAAAG ACAAAACAGA ACCCGTGGGC CCAGCTCGGA 540  
 50 AGGTGCGTGG AGAAGGCTCC GACGTCTCCG AAGTGCAGCC CTTGGGATGG CATTCCGTTG 600  
 TGTGCCTTAT TCCTGGAGAA TCTGTATACG GCTCGCCTAT AAGAAATATA GCCTCTTCAT 660  
 55 GCTGTATTAA AAGGACTTTT AAAAGCAAAA AAAAAAAAAA AAAAATCGA GGGGGGGCCC 720  
 GGTACCCAAT TCGCCCAATA GTGAGTCGTA TTACAATTCA CTGGGCCGTC STTTTAACAA 780  
 CGTCGTGAAC TGGGAAAACC CTGGCGTTTA CCCAACTTAA TCGCCTTGCA GCACATCCCC 840  
 60

CTTTGCCCAG CTGGCGTTAA TAGCGAAAAA NGCCCGCACC CGAATCGCCC TTCCCAACAG 900  
 TTTGCGCAGC CCTGAATGGC GAAATGGCAA ATTGTAAGCG TTAAATATTT TKKTAAAAAT 960  
 5 TCCNCGTTWA AWTTTTTGTT TAAATCARCT CAATTTTTTT AACCCAATAA GSCCGAAATC 1020  
 CGGCAAATCC CCYTTATTAA TTCCAAAAAA ATAAACCSAA AAWGGGTTTG AATTTTTTKT 1080  
 10 TTCCCCAYTT TTGGAACAA AWTYCCCCCT TTTTAAAAAA GTTGAACCC CCAMCCYTCC 1140  
 AAAGGGGAAA AAACSYTTTT YTGCGGGGNA ANGGGGCCCC CMTACTTTNA ACAYCCCCCC 1200  
 CCAAWCAATT TTTTGGGGG GTCCCNAAAG GTCCCCCTAA AANCTTTTTT CGGAACCCNA 1260  
 15 AGGGGANCCC CCCATTTAAA ATTTTNGGTN 1290

20 (2) INFORMATION FOR SEQ ID NO: 197:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1020 base pairs  
 (B) TYPE: nucleic acid  
 25 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 197:

30 GGTGTCCCTG GATGTCGTG TAGGTGAGTT TTACCAAGGA TTATGGTAAC AAATGAGTGA 60  
 GACCTCTATG GAGAAAATAT TGAAGNNCAT TAAAGAAGAC CTCATANTAG GAGAGAATGT 120  
 35 SCTTTGGAGG ATTTGTATTG AGCTTTTACA GTATTCATTT TTCAACTCAA GGCAATGGCT 180  
 TTCTACACCA ACTCTAATCC ATAAACGGGT CTTATGACAT CTATGAAGTA GTAGCAAGAC 240  
 ATGCTTAGTG TGTATTCTC TCTTTGAGAC ACTGTAATTT CTACCAGAAA TTTCCAGAGC 300  
 40 ATTATGTAGG TAGAAAAAAA TGCAAGCAAG CTGTTAAAGA TCTTGGATCC CATTATATAG 360  
 TATGTATAGC TGAAATCTGT AATCAATCA CTTTCTCTCT TTTATCCTCT AACCAAAAAA 420  
 45 TTGTTAATT TTGCATCCCA AATGTTTTTA ATCTTTGTAT ATTTTTTAAA AAYCCTTTTC 480  
 TCCTCATCAT TGCCTTTTTT GTGGTTGFAA ATAGACTTAC TTGCACTTG AAGATGAGTT 540  
 ACTCCTTGTC ATCTTACAAA TATGTGATAT GGTAATTTTC ATAACAGATG TCAGTTTTGA 600  
 50 ACCAAGAATT GGTGATTTGT TTATAAGAAA AAAACTGGCT TCATTCTGT GAAATTGCTC 660  
 TTTGAAAATT TCTTTTACA CGTGAAGCC AACTGAGATA CCGTGATGGT GTTGATTTCT 720  
 55 TTCAATGATG CTTACCATCT ATTTAGCCA CTGAGCCTTT TATTATTGT CTATTGTAA 780  
 AGTTTATTIG TCTTAACTCA TTTAATAAAT ATACTGTTTA TCTGTTCTG AATGGGGACT 840  
 GAACTTTTG GATATTGATA TTGATTTGAA AATATTTTGG AATTTTTTCT ACTTGAAATT 900  
 60 TTAGAAATCT AATGAAAAAT TCTATAATGT ACTGAAAGTA WGGTTGTGTA CAGTGAKCAC 960

TCTCTAATAA TATGATGNCT TGCCCTAAAN GAGGNGGGAC ATGTCCCACT TTCCACCACG 1020

5

(2) INFORMATION FOR SEQ ID NO: 198:

10

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 524 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 198:

AATTCCCGAA GCTGAGGGTT GTGTGCCNTC GGGCGAGCCA AGTCTTTTGA CCGGACCCCTT 60  
CCCGCGCAG AAGANCTGAA GTTGATTGA GAGCCTGRTT TGGGGITRA GCCGAGCTGC 120  
TGCGGGCTTY GTCCCGGCC AGGACACAAG YTAATTGCAA CGGGCGGCG CCTGGCTTAT 180  
GATGTTCTTC AACCCAGGGG CGCCTCTGC CCTCTACTCG TGCCAGGCC ACTTGCCAGG 240  
CAGGAGCCCT CCCAAGCCT TCAGGGCTGC TCGGAGTCAC CTGTTGGAAT GGAATAAAG 300  
GACCCCTGTG TGGGAACAGG TGCTCAAAC ACCCTGCTGC TGGCTGCCAG GCAGGCCCTC 360  
TGGAAGGGAA GGGGCAGGAC TCATCAGGAC CTCCTGGAC CCTGCAGGGC AGGCAGTTGG 420  
CCCAGCCCA AGCATTTGGC TCTGCTTGGC CCAAGGGGAC AGGAAGCCTC TTGGGCCTCT 480  
TCCCTTCTG GACAAGGCC CCTGCCITTG CCTCACATAA ACTG 524

35

(2) INFORMATION FOR SEQ ID NO: 199:

40

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 332 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 199:

GTGATACAAG GAAGGGTGAT CATCATCTGT CACCATGCAA TTCCTGCTCA CAGCCTTTCT 60  
GTGGTGCCA CTTCGGCTC TTTGTGATGT CCCCATATCC CTAGGCTTCT CCCCCTCCTA 120  
GAAGGGCTTC TTGATAGATT AGAAAATAAG AATGAGTGAC ATTTCCTATG TGCATATAAG 180  
AAGGAGCCAC AAGACATGTC TTTTAAATAA AAGGACAGTG TCCATCCTTT TAGCTGCCGA 240  
ATAGAACCTT GGTCTCATCC TCCTGGAGCT AGGSCITAAA ACAGCTTCTG TGTTCCTSAT 300  
TKGTCTCART GTTTTGCCAA GGTTTATTC GG 332

60

## (2) INFORMATION FOR SEQ ID NO: 200:

- 5 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 376 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
10 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 200:

CCAGGGAAGC CCCARGCCTG TCCTGAATTG ACATCAGTGC TTCCCTGAAC TGCCTCCCCC 60  
15 ACCCCTGGGC ATTATCCCAG GAAACTTATG TTTTCTAGAA GCTAAGCAGC TGCTGGGACT 120  
CAGGGACTGG TGCAGGTAGG CTGAGTGGCA GCTCAGTCCT AGAAGGTCTC TGAAGATCTG 180  
GACTGAGGAC CYTGCTACTC CCCAAGCCAG AGCCCATCAG CCAGGCCTGC TGTGAGCCAC 240  
20 CTGCCTGTGG AGTGCTGAGC TCAACCAAAG GCTGGCAAGC TCTGGGCCTC ATTTAAGGGA 300  
TTCTGATGAG CCGATGGGCC CTGGAGGCAG CCCATTAAAG CATCTGGCTC GTTTTGGAA 360  
25 AAAAAAAAAA AAAAAG 376

## (2) INFORMATION FOR SEQ ID NO: 201:

- 30 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1192 base pairs  
35 (B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 201:

40 CCCAGTATAT TTCTATAACA TTTATTTTAG TGAAC TTATA ATGTTTCTTT GTATTAAATT 60  
ATTAGATTAT ATCTTTAGAT AATATTGTTA CTNAATTAGT AGGTAATATA TATTTTATTC 120  
AAAAATAAAT TGTCATCTA ATGTCTACCA ATTAATGTAC TTGTAGATGT ATCTTATCTT 180  
45 AACTTGAGTC TTTGCTGCCC CTAATGAGGT GTGAAGGACT CTCTCCCCCT GGGGAAGTTT 240  
TTCTTTTCA GGAGGGAGGA GGGCTTTCCC AGGTAATGTG TCTAGAGTGT TGGGCAGAAR 300  
50 AATCTGGGAC CACACCACAC CAGTCTCTC CTTAATCCAC GTCATTTGCC TTCTATCCCA 360  
GCTATGTTTC CAGTGTCTC TGGGTGTTTC CAAGAGCAAC AAGAAAYGAA TAAATCTCTG 420  
KTGAGTTGTT TATTTGTTCT TCACTTTGTT TTAACTGTA WTTTCTGAGT TTATGGGTGT 480  
55 CTGTGAATTA AAAAGGAAAA GTRGAAATAA GTAAACTCA GGTGAAGGA AATATACATA 540  
AATAAGATAA AGCTGACCTG TAGATATARR CAGGTTATAA RAGCTTAGAG TTGTCTAAGT 600  
60 TGRGTGCAAA KTTTCCTCTG ATCTTTCTGA TGCCGARACA AAAAAGGCAG TCATGTTTGT 660

5 WATGTGATTG GAATGGAACC CGARAAGAGA GCAYGCTGTG TTCTTGGGGA CAGGAAAGCT 720  
 TGYGTGCACC AAGTCTKAAC CACCACCTTC ATGGGACATA GRTTATGTGC TGGAACATAT 780  
 TTCACACCGG CCTGGCAGTA AACACTTGTA GTGTTGTGCA GTGGAAACGG TCATCTTCCG 840  
 CTAAGCACG GCGTGTGTG CAGCGGAAAT GGTCATCTGC TGCTAAAACA CAGCTTCCAT 900  
 10 CGTAATGTAT GCTCCTTACT CAAAGAGTGT GGTCCCAAAC AGCCTTTGGG AGGTCCCTCT 960  
 TGATTCATGG ATGAAACCTG GAACATCTTG AGGACTGAGT TAACCATAGG TCCTTAAATA 1020  
 15 ACTCTCCACA CGTTTTCTTT AGTTTATCTC TACATGCAGG GTGTGCAGCA GCCTGTTCAA 1080  
 AGTCATATTT TCTGGGAAAT ATTTCCAGTG TTTATTTGCA CTTAGCCCA CTCTGTGTAG 1140  
 CCTTATTTCT TCTAAACTCA CCATTAATCT GAATAATAGT CAAATTTAGG GG 1192

20

## (2) INFORMATION FOR SEQ ID NO: 202:

25

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 589 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

30

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 202:

ATCTTGGGCT ATCTTTGACA GGGGATCTTT GCAAGTTGAT GCITTCTACA AGTGAATATA 60  
 35 GTCAGTCCCC AAAGATGGAG AGCTTGAGTT CTCACAGAAT TGATGAAGAT GGAGAAAACA 120  
 CACAGATTGA GGATACGGAA CCCATGTCTC CAGTTCTCAA TTCTAAATTT GTTCCTGCTG 180  
 40 AAAATGATAG TATCCTGATG AATCCAGCAC AGGATGGTGA AGTACAACTG AGTCAGAATG 240  
 ATGACAAAAC AAAGGGAGAT GATACAGACA CCMGGGATGA CATTAGTATT TTAGCCACTG 300  
 GTTGCAAGGG CAGAGAAGAA ACGGTAGCAG AAGATGTTTG TATTGATCTC ACTTGTGATT 360  
 45 CGGGGAGTCA GGCAGTTCCG TCACCAGCTA CTCGATCTGA GGCACCTTCT AGTGTGTTAG 420  
 ATCAGGAGGA AGCTATGGAA ATTAAAGAAC ACCATCCAGA GGAGGGTCTT TCAGGGTCTG 480  
 50 AGGTGGAAGA AATCCCTGAG ACACCTTGTG AAAGTCAAGG AGAGGAACTC AAAGAAGAAA 540  
 ATATGGAGAG TGTTCCGTTG CACCTTTCTC TGACTGAAAC TCAGTCCCA 589

55

## (2) INFORMATION FOR SEQ ID NO: 203:

60

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 847 base pairs
  - (B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 203:

5 GGACAGAGCG CAAGCTGCTG GCCGCCATCA ACGCGTTCCG CCAGGTGCGG CTGAAACACC 60  
 GGAAGCTCCG GGAACAAGTG AACTCCATGG TGGACATCTC CAAGATGCAC ATGATCCTGT 120  
 10 ATGACCTGCA GCAGAATCTG AGCAGCTCAC ACCGGGCCCT GGAGAAACAG ATTGACACGC 180  
 TGGCGGGGAA GCTGGATGCC CTGACTGAGC TGCTTAGCAC TGCCCTGGGG CCGAGCAGCT 240  
 TCCAGAAGCC AGCCAGCAGT CCAAGTAGCT GGACCCACGA GGAGGAACCA GGCTACTTTC 300  
 15 CCCAGTACTG AGTGGTGGAC ATCGTCTCTG CCACTCCTGA CCAGCCTGAA CAAAGCACCT 360  
 CAAGTCAAG GACCAAAGGG GGCCTGGCTT GGATGGGTTG GCTTGCTGAT GGCTGCTGGA 420  
 20 GGGGACGCTG GCTAAAGTGG GGAGGCCTTG GCCACCTGA GGCCCCAGGT GGGAACATGG 480  
 TCACCCCCAC TCTGCATACC CTCATCAAAA ACACTCTCAC TATGCTGCTA TGGACGACCT 540  
 CCAGCTCTCA GTTACAAGTG CAGGCGACTG GAGGCAGGAC TCTTGGGTCC CTGGGAAAGA 600  
 25 GGGTACTAGG GGCCCGGATC CAGGATCTG GGAGGCTTCA GTTACCGCTG GCCGAGCTGA 660  
 AGAACTGGGT ATGAGGCTGG GCGGGGCTG GAGGTGGCGC CCCCTGGTGG GACAACAAG 720  
 30 AGGACACCAT TTTTCCAGAG CTGCAGAGAG CACCTGGTGG GGAGGAAGAA GTGTAAGTCA 780  
 CCAGCCTCTG CTCTTATCTT TGTAAATAAT GTTAAAGCCA GAAAAAAAAA AAAAAAAAAA 840  
 35 AAAAAAA 847

(2) INFORMATION FOR SEQ ID NO: 204:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 852 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 204:

50 ACAAACATAC TCGCAGGAAG GAGTCTCATG CTGCCCGCAG CATCAGCGCA ACNCNTGGCC 60  
 GCCATCAACG CGTTCGGCCA GGTGCGGCTG AAACACCGGA AGCTCCGGGA ACAAGTGAAC 120  
 TCCATGGTGG ACATCTCCAA GATGCATATG ATCCTGTATG ACCTGCAGCA GAATCTGAGC 180  
 55 AGCTCACACC GGGCCCTGGA GAAACAGATT GACACGCTGG CGGGGAAGCT GGATGCCCTG 240  
 ACTGAGCTGC TTAGCACTGC CCTGGGGCCG AGGCAGCTTC CAGAACCAG CCAGCAGTCC 300  
 60 AAGTAGCTGG ACCCAGGNAG GAGGAACCAG GCTACTTTCC CCAGTACTGA GGTGGTGGAC 360

ATNCGTCTCT TGCCACTCCN TGNACCCAGC CCTGAACAAA GCACCTCAAG TGCAAGGACC 420  
 AAAGGGGGCC CTGGCTTGA GTGGGTGGC TTGCTGATGG CTGCTGGAGG GGACGCTGGC 480  
 5 TAAAGTGGGK AGGCCTTGGC CCACCTGAGG CCCCAGGTGG GAACATGGTC ACCCCCACTC 540  
 TGCATACCCT CATCAAAAAC ACTCTCACTA TGCTGCTATG GACGACCTCC AGCTCTCAGT 600  
 10 TACAAGTGCA GCGCACTGGA GGCAGGACTC CTGGGTCCCT GGGAAAGAGG GTA CTAGGGG 660  
 CCGGATCCA GGATTCTGGG AGGCTTCAGT TACCGCTGGC CGAGCTGAAG AACTGGGTAT 720  
 GAGGCTGGGG CGGGGCTGGA GGTGGCGCCC CCTGGTGGGA CAACAAAGAG GACACCATT 780  
 15 TTCCAGAGCT GCAGAGAGCA CCTGGTGGGG AGGAAGAAGT GTA ACTCACC AGCCTCTGCT 840  
 CTTATCTTTG TA 852

20

(2) INFORMATION FOR SEQ ID NO: 205:

25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1354 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 205:

GATTGGGCAC GAGGCTTGCT GGAGCAGGAG AAGTCTCTRG CCGGCTGGGC ACTGGTGCTG 60  
 GCASGARCTG GCATTGGACT CATGGTGCTG CATGCAGAGA TGCTGTGGTT CGGGGGGTGC 120  
 35 TCGGCTGTCA ATGCCACTGG GCACCTTCA GACACACTTT GGCTGATCCC CATCACATTC 180  
 CTGACCATCG GCTATGGTGA CGTGGTGCCG GGCACCATGT GGGGCAAGAT CGTYTGCCCTG 240  
 40 TGCCTGGAG TCATGGGTGT CTGCTGCACA GCCCTGCTGG TGCCCGTGGT GGCCCGGAAG 300  
 CTGGAGTTTA ACAAGGCAGA GAAGCACGTG CACAACCTCA TGATGGATAT CCAGTATACC 360  
 AAAGAGATGA AGGAGTCCGC TGCCCGAGTG CTACAAGAAG CCTGGATGTT CTACAAACAT 420  
 45 ACTCGCAGGA AGGAGTCTCA TGCTGCCCGC AGGCATCAGC GCAANCTGCT GGCCGCCATC 480  
 AACGCGTTCC GCCAGGTGCG GCTGAAACAC CGGAAGCTCC GGAACAAGT GAACTCCATG 540  
 50 GTGGACATCT CCAAGATGCA CATGATCCTG TATGACCTGC AGCAGAACTT GAGCAGCTCA 600  
 CACCGGGCCC TGGAGAAACA GATTGACACG CTGGCGGGGA AGCTGGATGC CCTGACTGAG 660  
 CTGCTTAGCA CTGCCCTGGG GCCGAGGCAG CTTCAGAAC CCAGCCAGCA GTCCAAGTAG 720  
 55 CTGGACCCAC GAGGAGGAAC CAGGCTACTT TCCCCAGTAC TGAGGTGGTG GACATCGTCT 780  
 CTGCCACTCC TGANCCAGC CCTGAACAAA GCACCTCAAG TGCAAGGACC AAAGGGGGCC 840  
 60 CTGGCTTGA GTGGGTGGC TTGCTGATGG CTGCTGGAGG GGACGCTGGC TAAAGTGGGK 900

5 AGGCCTTGGC CCACCTGAGG CCCAGGTGG GAACATGGTC ACCCCCACTC TGCATACCCT 960  
 CATCAAAAAC ACTCTCACTA TGCTGCTATG GACGACCTCC AGCTCTCAGT TACAAGTGCA 1020  
 GGGGACTGGA GGCAGGACTC YTGCGTCCCT GGGAAAGAGG GYACTAGGGG CCCGGATCCA 1080  
 GGATTCTGGG AGGCTTCAGT TACCGCTGGC CGAGCTGAAG AACTGGGTAT GAGGCTGGGG 1140  
 10 CGGGGCTGGA GGTGGCGCCC CCTGGTGGGA CAACAAAGAG GACACCATTT TTCCAGAGCT 1200  
 GCAGAGAGCA CCTGGTGGGG AGGAAGAAGT GTAACTCACC AGCCTCTGCT CTTATCTTTG 1260  
 15 TAATAAATGT TAAAGCCAGA AAAAAATAAA AAAAAAAAAA AAAAAACTCG AGGGGGGCCC 1320  
 AGACCCAATC TCCCTATAGT AAGNCGCCNN ANAN 1354

20

(2) INFORMATION FOR SEQ ID NO: 206:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1378 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 206:

30 TCCCCAGGTG CACAGCCAGG GCCCTCCTGT CTGCAGGAGA ATTCACAGCT GGTGTGGGAC 60  
 TCAGCCCTTA GNCATTCAA AGCCTTAATG TTGTAATCAT ATCTTACGTG TTGAAGACCT 120  
 35 GACTGGAGAA ACAAATGTG CAATAACGYG AATTTTATCT TAGAGATCTG TGCAGCCTAT 180  
 TTCTGTCACA AAAGTTATAT TGTCTAATAA GAGAAGTCTT AATGGCCTCT GTGAATAATG 240  
 40 TAATCCAGT TACACGTGA CTTTAAATAG CATAAGTGA TTTGATGAAA GGACGTCAAA 300  
 CAATGTGGCG ATGTCGTGGA AAGTTATCTT TCCCGCTCTT TGCTGTGGTC ATTGTGTCTT 360  
 GCAGAAAGGA TGGCCCTGAT GCAGCAGCAG CGCCAGCTGT ANATAAAAAA TAATTCACAC 420  
 45 TATCAGACTA GCAAGGCACT AGAACTGGAA AAGACCACAG AAAACAAAGA ATCCAACCTT 480  
 TTCATCTTAC AGGTGAACAA ACTGTGATGA TGCACATGTA TGTGTTTGT AAGCTGTGAG 540  
 CACCGTAACA AAATGTAAAT TTGCCATTAT TAGGAAGTGC TGGTGGCAGT GAAGAAGCAC 600  
 50 CCAGGCCACT TGACTCCAG TCTGGTGCCC TGTCTACACC AGACAACACA GGAGCTGGGT 660  
 CAGATTCCCC TCAGCTGCTT AACAAAGTTC CTCGAACAGA AAGTGCTTAC AAAGCTGCCT 720  
 55 TCTCGGATAC TGAAAGGTCG AGTTTCTGA ACTGCACTGA TTTTATGCA GTTGAAAAAA 780  
 AAAAAAGCT ATTCCAAGA TTTCAAGCTG TTCTGAGACA TCTTCTGATG GCTTTACTTC 840  
 60 CTGAGAGGCA ATGTTTTTAC TTTATGCATA ATTCATTGTT GCCAAGGAAT AAAGTGAAGA 900



AACAGCACCT TTTAATATAT AGGTCTCTCT GGAAGAGACC TAAATTAGAA AGAGAAAAC 960  
GTGACAATT TATATCTC ATTCTTAAAA AACACTAATC TTAAC TAACA AAAGTTCTTT 1020  
5 TGAGAATAAG TTACACACAA TGGCCACAGC AGTTTGTCTT TAATAGTATA GTGCCTATAC 1080  
TCATGTAATC GGTACTCAC TACTGCCCTT AAAAAAAAA ACCAGCATAT TTATTGAAAA 1140  
10 CATGAGACAG GATTATAGTG CCTTAACCGA TATATTTTGT GACTTAAAA ATACATTTAA 1200  
AACTGCTCTT CTGCTCTAGT ACCATGCTTA GTGCAATGA TTATTCTAT GTACAACTGA 1260  
TGCTGTCTT TATTTTAATA AATTTATCAG AGTGAAAAA AAAAAAAAA AAAAAAAAA 1320  
15 AAAAAAAAA AAAAAAAAA AAAAAAAAA AAAAAAAAA AAAAAAGAA NAAAAANA 1378

20 (2) INFORMATION FOR SEQ ID NO: 207:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1166 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 207:

30 AANCCACTGC ANTTTAAACC CCTCCCTC CAAGAAAGTT CACAACCGC CATGGATGAC 60  
CCTCATTTTA GATGGGCCNC AATATTTAAG ATGGACTGRG GMCCCARAG ACTGACCCTT 120  
35 GAAAGGGGA CTCAGAAGAA AGATCCTTGA CATTGCCMAA CATGCTGGC TTGTCCAACA 180  
CAGTGATGCG GTCATCGAG AARCGGCTT TCCMAGGACA AGTACTTTAT GATAGGTGGG 240  
ATGCTGCTGA CCTGTGTGT CATGTCCTC GTGGTGCACT ACCTGACATG AGCCAGCCAC 300  
40 GCTCAGTGGC TGAACAGCAT TCCACAGCC TGCAAGTGTG TGTGTGTGTG AAAGAGAGAG 360  
GGGGCCAGA GGCGCCTTT TGAAATGTTT GCCTGTCTGA ACTGTGAAGA CACTTGGGAG 420  
TGATTGTGGT CTAATTTCCA ACCTGCTCTG TTTTCTGTGA CATCTTGAG GGGAGCTAG 480  
45 TGCCAMCACC ATGCGCGGTG CTTAGGAAAT GAAAGAAGTC CCGGTCTGT CTCTCTCACT 540  
CTGCTCTCA MTGGGGGAGG GAAAGAATGG CTTTGGTGGC TTTGTTTACA CAGCTGATGC 600  
50 GTGSCCTGGG AAGGTGTCCA CAGTGAGCCC TGTGTGCAGG ACTGTCCACN ACGGTTTACA 660  
CCTTGTCACT ATCAGGCCTT TCTGCTCCT GATAGGTGG AGCAAAAGTG GAAAGGAAAG 720  
GAAAGAGGCY TTTTCTTACA GCCATTATAT TAAATAGTAG GTCGATTAC ATCYTCGTGC 780  
55 TCCTGGCCAC CCTCCCTGT GCCTCAGTGA CATGTAGATG ACTGACTGCC AATACTGTG 840  
ACCATTCCT GGAAGCAGCT ACCTAGGGGA AACAAGATGT AGTGCTATTG CCGATAACAA 900  
60 GTAAGATTTT CCACACTACA GCTGGGTGTT TCTCTTTTCT AAAGTGAGGC CAGTGTATT 960

5 TCCCCGGAGT GTTCAGTCTT GACCCTAGTC ACTGATTTT TCTAGTTGTT AATAGAGTGG 1020  
TTGGGCTTTT AAGGTTTACA GACTGTGGGC TTGGGCACCT GCGCCACAGG STTTTGTGGG 1080  
GGCCTTTGCC CCTTAGRAAA GTAGCTTTTA GGGGCAAAGA TTTGTTGATT TTCCCCATTA 1140  
CAGTCTTCAG CTCNAGGGTT TTAATA 1166

10

(2) INFORMATION FOR SEQ ID NO: 208:

15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 697 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 208:

TACTTCTAGG ATTATAAGGA ATTAACATTG AGATGACATT TCCATTTGAG AAGGAAAATA 60  
25 GTTGCTTTCA GTGCCTTTTA TTTGATTCTT GGAGAGAGCA GACTCGCACS AACATTCAAC 120  
CCCAGCGCTG ATATGACAGT AATCCTCAGA GGCAGAGCCC AGCACAAAAC AGCAATGCTA 180  
GAAAGTTACA ATTGGAAAGT TTCCTGCCAG CTTCGGGAAT GACACTGCAA AGCTGATGCC 240  
30 AGAAACTGCC AGRGTAATTC TCCTCATTAC TGCTCTACCC ACCCACTTTC AGCTCCCCAA 300  
ATTAAGTAGT GCAGTTGACT AATTCTCTTT ACCTTTTATCA TTTARGGTGA RGCATTGCAC 360  
35 AAAAAGTCTC GACTTTGCCA TATAAGGGCT GTGGTTCTCT GTGTCCCTT GGATAAGAGG 420  
CATCACCATT ATCTGGAAC ATGCAGTAAA TGCAGATTNT TCATCTTCTC CCCAGACCTC 480  
CTGAGTTAGA AATTACACAAG TTCTCCAGGT GATCTCATACT ATGCTAAAGT TTGAGAACCA 540  
40 TTGAGTAAAG TTAATGCATT AAGAAGAGAT TAGATAGGGA TGGTGGCGTA TCTTCCTACA 600  
GTTTCCCTGT TAACAAGAAA GTCAGAGGTC AGTTGATCAG ACATTAGATT ATTTATTGCT 660  
45 AAAACTAAAA AAAATTAAAA AAACTGGAG GGGGGCC 697

50

(2) INFORMATION FOR SEQ ID NO: 209:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 932 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 209:

60

CGTGAGTCAC CTCTCTATAG TGGGCGTGGC CGAGGCCGGG GTGACCCTGC CGAAGCCTCC 60

	GCTGCCAGAA ACCATGTTCA AGGTAATTAA AAGGTCCGTG GGGCCAGCCA GCCTGAGCTT	120
5	GCTCACCTTC AAAGTCTATG CAGCACCAAA AAAGGACTCA CCTCCCAAAA ATTCCGTGAA	180
	GGTGTATGAG CTTTCACTCT ACTCAGTTCC TGAGGGTCAA TCGAAGTATG TGGAGGAGGC	240
	AAGGAGCCAG CTTGAAGAAA GCATCTCACA GCTCCGACAC TATTGCGAGC CATAACAAC	300
10	CTGGTGTGAG GAAACGTACT CCCAACTAA GCCCAAGATG CAAAGTTTGG TTCAATGGGG	360
	GTTAGACAGC TATGACTATC TCCAAAATGC ACCTCCTGGA TTTTTCCTGA GACTTGGTGT	420
15	TATTGGTTTT GCTGGCCTTA TTGGACTCCT TTTGGCTAGA GGTTCAAAAA TAAAGAAGCT	480
	AGTGTATCCG CCTGGTTTCA TGGGATTAGC TGCTCCCTC TATTATCCAC AACAGCCAT	540
	CGTGTGTC CAGGTCAGTG GGGAGAGATT ATATGACTGG GGTTTACGAG GATATATAGT	600
20	CATAGAAGAT TTGTGGAAGG AGAACTTTCA AAAGCCAGGA AATGTGAAGA ATTACCTGG	660
	AACTAAGTAG AAAACTYCAT GYTCTGCCAT CTTAATCAGT TATRGGTAAA CATTGGAAC	720
25	TCCATAGAAT AAATCAGTAT TTCTACAGAA AAATGGCATA GAAGTCAGTA TTGAATGTAT	780
	TAAATGGCT TTCTTCTTCA GGAAAACTA GACCAGACCT CTGTTATCTT CTGTGAAATC	840
	ATCCTACAAG CAAACTAACC TGAATCCCT TCACCTAGAG ATAATGTACA AGCCTTAGAA	900
30	CTCCTCATT CTAATGTGCT ATTTATGTAC CT	932

35 (2) INFORMATION FOR SEQ ID NO: 210:

(i) SEQUENCE CHARACTERISTICS:

- 40 (A) LENGTH: 661 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 210:

45	GTCAATCTTT AAATAAAAGC TTTCTGTGTT AAAGCTTTTC AAAGGAGCAG ACCACCTTGA	60
	AGATCCCCC TAGGGTGTAT ATGTGTCTAA TTCATTTTAT AAAAATTATT CTGTCTTCA	120
50	TTTTAAAGCT TTGGCTATAT AGTCAGAAAT GTCCTAAATA ACAAACTATT TGTATTTAA	180
	TTTAGGGAAG ACTAAAGGA AGAAAAATGA AAATCAGTC TTTATGTAAG CTCCAAGGAT	240
	ATTAGGGCTT AAAGGGCTTT TCTAGTTTGA TGAGAATTG TACTACTGAT TTTTATATAT	300
55	TCCTGTTTTT GAGATGAACA GATCTCTGGG GAAATGTTG AGTTACAATG GCATTTCACT	360
	GTGATCCCTC TCAAGCTCAG ATCAGTTCTA TAACCAATG ACAACCTGTC TCTTTGGTTT	420
60	ACTGTCCTGT GAAATGTCAG CTCAGTTTC CCAGAAGTCG TGTGTTTATG ATGAGTCAGA	480

GTGCTTTTCC TCGGTGGGAC AGTTGCTGGC CCTCTTAATT TTGGTGTATG TGCTTCCAAG 540  
TATCTAAACC TCCAGTCTGA TCTGTATATG CTATCCTAAC TGTTAATTGT ATTATTGATT 600  
5 ATGTTGATTA TCTTGCTTGA AGGTTCATAC TTTTCAATTT GATAGAAATA AAGTTTTTTT 660  
C 661

10

(2) INFORMATION FOR SEQ ID NO: 211:

(i) SEQUENCE CHARACTERISTICS:  
15 (A) LENGTH: 592 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 211:

GAAACTGACA TTGTTAAACA CACTAAAACA GAAGTACTTA CCTCTGAAG ATTTAATATA 60  
TAATGGTTGA CATGATACAT GTACATGAAT GGAATGACCA GATGCTTATG GTCTACATTT 120  
25 TCCTTTATCC TGTTAGTATT ACCTTCCTTA ATCTTTGTTC CTTAACATGC TAAATTCCTC 180  
TTCAGTGTTC ATTTTCTAGT GACAGAATGC TAACATTTCT TACACCCTGG CAGAAGGGAG 240  
30 AGAAATGTGT TTTGGGGTGG GTAACTAAAT TTTTGAGTGA AATATCATAA GATGAGAATG 300  
GAAAGAGGGA GACACAAAGA GTTATAACAA AAAACAATG GTTTTTTTAG CCATTTGACT 360  
GGCTCTTTAA ATAGTCTACA AGACATTCAC GTTNAACATC ACTTTTAGTG AAATAAAATG 420  
35 TGCCATACTA GTATGTGCTT CAAAAGGGCA AATGTGCTTT AGTGCCCTAA GGCTAAATTT 480  
TGGTCATTTG ACATCAGAGA TGTGTAAAGT ATTGCACTTA ATACGCACCT ATTTCTCAAT 540  
40 AGTGNTATTT TTTTGGCTAG CATTTNCTTT ACCACTAACC TTGTTGGATA GC 592

45 (2) INFORMATION FOR SEQ ID NO: 212:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 938 base pairs  
50 (B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 212:

55 TGGAGTGGCT TTCCAGCTGA ATGAATCCTA TGTCTGCGT GCAGGTGGTT GGTTTTCAAT 60  
GTCTTTSCTA ATTTTMTTCC TATGGCTCT TGGGAGTTN CTTTGTTC TCCTGTGTTT 120  
GCCAGCTTT AATAAAACCA GCGCAAACA AAAACCATAG CATTCTGAAA CAATAGGGGG 180  
60

	CCCACATGG ACCCAGTATG TCACTTTAAT GGAATTCAAG AAAAAATCTG AATGGGAAAA	240
	TGACACTAGG AATGTATACT CCACACATTT TATGCCATAT AATGGTGTGT TTTCTTAATT	300
5	TTGTTTCTTG TGGCGAAATG TGGCTTTCAA ATTAAAATGM CCTTTTCTTC TTKGAACTT	360
	TTTGTTTKGA CTKGTATAAT TAAGGGTTTG GAAAGATTCA TAATTMIGAG AGAGGTTTGC	420
10	AACCAGGAGA TACAAAGAAG TCTCAGTAGT AATCTTGTTT ATGTGCTTTT ACAGCCAGCT	480
	ACATTTAAGR ATGTATTAGT TACAGAAATT ATATGTCTGT GTATGTGTCT CTAICTAATA	540
	AAGTACATGC CTCACATAA TGCGGTGCTG TCCATCTCGG CAAATACTGG CCAAGTCCCT	600
15	TTATGACAGG CACACAGAAA CCATAGCATG GTCTGGCTTT CAGAAAATGC CTCTCATCTT	660
	TCCTGGAACC TTATTTTGCT AAATGTCTGT TTTCTTGTA TTTGTTGTAC CTCACAGCAC	720
20	CATTGTGACC ATGGTGATGC CTCATTGCA TGATATGTAC CTTGTGTTTA ATGTGAAATA	780
	CATTTTCATT GAAGAGTCTG ATGACTTGCT AGCGTTTAT TTTTCTGTA AGCTCAATGT	840
	GCTGAAACCA AACCAGGCTT TTA AAAACCT GTGTAGAAGA AAACCAAAAA ATCCTGTGTG	900
25	GGTGTCTTT CCTGTCAAA CTCATTAAAA ATTCTTT	938

30 (2) INFORMATION FOR SEQ ID NO: 213:

(i) SEQUENCE CHARACTERISTICS:

- 35 (A) LENGTH: 1079 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 213:

40	AGCCTGCCGG GAGAGTGGT GCATCTRARA GGCTGGTCTG GACTGTGTT TGGGGGAGGT	60
	GGGAGCTGTT TTAACCGTGT GCGCCCTCTC CTGTGCKGC GTGGGCATCC CCCGGGGCAG	120
45	TGGAACGCGG GCGTCTCTCC AGCTTCCGAG TCCAGCCAGC CTGGGCGCGG GCGCGCCCC	180
	CGAGACACCC GAGGAGTCCG TTCCTCCCTG GTTACGTGGA CTGTGGAGCT GGTCTCTTGT	240
	GGCTCAGCGC CGTGCGGAGG TTGAAGCGTA CCTGCGGAGG TCCACCAGG GCGGTGAGGA	300
50	GGAGGAGGAA GGCATGAGC CGAGCTGAG GAATCCGTGY TCCAACTCT AACTCAAGG	360
	RTGCMCTGCG CAACTCTGGT GCGCATGGC TGGGGCAGAT GTCTTGAG TTCTACCAGA	420
55	AGAAGAAGTC TCGCTGGCCA TTCTCAGACG AGTGCATCCC ATGGGAAGTG TGGACGGTCA	480
	AGGTGCATGT GGTAGCCCTG GCCACGAGC AGGAGCGGCA GATCTGCCG GAGAAGGTGG	540
	GTGAGAACT CTGCGAGAAG ATCATCAACA TCGTGGAGGT GATGAATCG CATGAGTACT	600
60	TGCCCAAGAT GCCACACAG TCGGAGGTG ATAACGTGTT TGACACAGC TTGCGGGAGC	660

5 TGCAGCCCTA CCTGTACAAG ATCTCCTTCC AGATCACTGA TGCCCTGGGC ACCTCAGTCA 720  
 CCACCACCAT GCGCAGGCTC ATCAAAGACA CCCTTGCCCT CTGAGCGTCG CTGGATCTCT 780  
 GGGAGCTCCT TGATGGCTCC CAGACCTTGG CTTTGGGAA TTGCACTTTT GGGCCTTTGG 840  
 GCTCTGGAAC CTGCTCTGGG TCATTGGTGA GACTTGGAAG GGGCAGCCCC CGCTGGCTTC 900  
 10 TTGGTTTGT GGTGCGCAGC CTCAGGTCAT CCTTTTAATC TTTGCTGACG GTTCAGTCTT 960  
 GCCTCTACTG TCTCTCCATA GCCTGGTGG GGTCCCCCTT CTTTCTCCAC TGTACAGAAG 1020  
 15 AGCCACCACT GGGATGGGA ATAAAGTTGA GAACATGAGT TTGGGCTGAA AAAAAAAAAA 1079

20 (2) INFORMATION FOR SEQ ID NO: 214:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3791 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 25 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 214:

30 TGAAGCAGGC GCTCTTGGCT CGGCGCGGCC CGCTGCAATC CGTGGAGGAA CGCGCCGCCG 60  
 AGCCACCATC ATGCCTGGGC ACTTACAGGA AGGCTTCGGC TCGTGGTCA CCAACCGATT 120  
 CGACCAGTTA TTTGACGACG AATCGGACCC CTTGAGGTG CTGAAGGCAG CAGAGAACAA 180  
 35 GAAAAAAGAA GCCGCGGGG GCGCGTGG GGGCCTGGG GCCAAGAGCG CATCAGGGCC 240  
 GCGGCCAGA CCAACTCCAA CGCGGCAGGC AAACAGCTGC GCAAGGAGTC CCAGAAAGAC 300  
 CGCAAGAACC CGCTGCCCCC CAGCGTGGC GTGGTTGACA AGAAAGAGGA GACGCAGCCG 360  
 40 CCCGTGGCGC TTTAAGAAAG AAGGAATAAG ACGAGTTGGA AGAAGACCTG ATCAACAAC 420  
 TCAGGGTGAA GGGAAAATAA TTGATAGAAG ACCAGAAAG CGACCACCTC GTGAACGAAG 480  
 45 ATTGAAAAG CCACTTGAAG AAAAGGGTGA AGGAGCGGAA TTTTCAGTTG ATAGACCGAT 540  
 TATTGACCGA CCTATTGAG GTCGTGGTGG TCTTGGAAGA GGTGAGGGG GCCGTGGACG 600  
 TGGAATGGGC CGAGGAGATG GATTGATTC TCGTGGCAA CGTGAATTG ATAGGCATAG 660  
 50 TGGAAGTGAT AGATCTTCTT TTTACATTA CAGTGGCCTG AAGCACGAGG ACAAACGTGG 720  
 AGGTAGCGGA TCTCACAAC GGGGAACGT CAAAGACGAA TTAACGACT TGGATCAATC 780  
 55 AAATGTGACT GAGGAAACAC CTGAAGGTGA AGAACATCAT CCAGTGGCAG AACTGAAAA 840  
 TAAGGAGAA GAAGTTGAAG AGGTAAAAGA GGAGGTCCA AAAGAGATGA CTTTGGATGA 900  
 GTGGAAGGCT ATTCAAAATA AGGACCGGC AAAAGTAGAA TTTAATATCC GAAAACCAAA 960  
 60

	TGAAGGTGCT GATGGGCAGT GGAAGAAGGG ATTTGTTCTT CATAAATCAA AGAGTGAAGA	1020
	GGCTCATGCT GAAGATTGGG TTATGGACCA TCATTTCGGG AAGCCAGCAA ATGATATAAC	1080
5	GTCTCAGCTG GAGATCAATT TTGGAGACCT TGGCCGCCA GGACGTGGCG GCAGGGGAGG	1140
	ACGAGGTGGA CGTGGGCGTG GTGGGCGCCC AAACCGTGGC AGCAGGACCG ACAAGTCAAG	1200
10	TGCTTCTGCT CCTGATGTGG ATGACCCAGA GGCATTCCCA GCTCTGGCTT AACTGGATGC	1260
	CATAAGACAA CCTGGTTC TTTGTGAACC CTCTGTTC AAGCTTTTGC ATGCTTAAGG	1320
	ATTCCAAACG ACTAAGAAAT TAAAAAATAA AAGACTGTCA TTCATACCAT TCACACCTAA	1380
15	AGACTGAATT TTATCTGTTT TAAAAATGAA CTCTCCCGC TACACAGAAG TAACAAATAT	1440
	GGTAGTCAGT TTTGTATTTA GAAATGTATT GGTAGCAGGG ATGTTTTCAT AATTTTCAGA	1500
20	GATTATGCAT TCTTCATGAA TACTTTTGTA TTGCTGCTTG CAAATATGCA TTTCCAAACT	1560
	TGAAATATAG GTGTGAACAG TGTGTACCAG TTTAAAGCTT TCACTTCATT TGTGTTTTTT	1620
	AATTAAGGAT TTAGAAGTTC CCCCAATTAC AAACGTGTTT TAAATATGG ACATACTGGT	1680
25	TTTAATACCT GCTTGCATA TTCACACATG GTCAACTGGG ACATGTTAAA CTTTGATTTG	1740
	TCAAATTTTA TGCTGTGTGG AATACTAACT ATATGTATTT TAACTTAGTT TTAATATTTT	1800
30	CATTTTGGG GAAAAATCTT TTTTCACTTC TCATGATAGC TGTATATAT ATATGCTAAA	1860
	TCTTTATATA CAGAAATATC AGTACTTGAA CAAATTCAAA GCACATTTGG TTTATTAACC	1920
	CTTGCTCCTT GCATGGCTCA TTAGGTTCAA ATTATACTG ATTTACATTT TCAGCTATAT	1980
35	TTACTTTTTA AATGCTTGAG TTTCCCATTT TAAAACTAA ACTAGACATC TTAATTGGTG	2040
	AAAGTTGTTT AAACACTTTA TTGTGTGTAG GCACATCGTG TCAAGTGAAG TAGTTTTATA	2100
40	GGTATGGGTT TTTTCTCCCC CTTCACCAGG GTGGGTGGAA TAAGTTGATT TGGCCAATGT	2160
	GTAATATTTA AACTGTCTG TAAAAAAGT GTCTGGCCAT TTGGTATGAT TTCTGTGTGT	2220
	GAAAGGTCCC AAAATCAAAA TGGTACATCC ATAATCAGCC ACCATTTAAC CCTTCCTTGT	2280
45	TCTAAAACAA AAACCAAAGG GCGCTGGTTG GTAGGGTGAG GTGGGGAGT ATTTTAATTT	2340
	TTGGAATTTG GGAAGCAGAC AGCTTTACTT TGTAAGGTTG GAACAGCAGC ACTATACATG	2400
50	AAATATAAAC CAAAAACCTT TACTGTTTCT AAATTTCTTA GATTGCTATT ATTTGGTTGT	2460
	AAGTTGAGTA TTCCACAGAA AGTGGTAATT ATCTCTTCTC TCTTCTCCA TTAGAAAATT	2520
	AGGTAAATAA TGGATTCTTA TAATGGGAGC ATCACCATT ATTAAAAAC ACATAGAATG	2580
55	ATGAATTAAA AAAGTTTCTT AGGATTGTCT TTTATTCTCC CACATTTATT GATAAACAGT	2640
	GAAGGAATTT TAAAAAATT TTTAAGAATT GTTTGTCAG TCATTTTATG AAATGTTCTA	2700
60	CCTGTATATG GTAATGTCCA GTTTTAAAAA TATTGGACAT CTTCAATCTT AAACATTTCT	2760

	ATTTAGCTGA TTGGTTCTCA CATATACTTC TAAAAGAAAC TTTTATGTTA TAAGAGTTAC	2820
	TTTTTGATA AGATTTAITA ATCTCAGTTA CCTACTATTC TGACATTTTA GGAAGGAGGT	2880
5	AATGTTTTTT AATGATGGAT AAAC TTGTGC TGGTGT TTG GATCTTATGA TGCTGAGCAT	2940
	GTTCGCACT GGTGCTAATG TCTAATATAA TTTTATATTT ACACACATAC GTGCTACCCA	3000
10	GAGATTAATT TAGTCCATAT GAACTATTGA CCCATTGTTC ATTGAGACAG CAACATACGC	3060
	ACTCCTAAAT CAGTGTGTTT AGACTTTTCA AGTATCTAAC TCATTTCCAA ACATGTACCA	3120
	TGTTTTATAA ACCTCTTGAT TTCCAGCAAC ATACTATAGA AAACACCTGC TACTCAAAAC	3180
15	ACAACTTCTC AGTGTCTATCC ATTGCTGTGC TGAGAGACAA CATAGCAATA TCTGGTATGT	3240
	TGCAAGCTTT CAAGATAGCC TGAAC TTAA AAGTTGGTGC ATTAGTTGTA TCTGATGGAT	3300
20	ATAAATTGTC CTCCTAGTTC ACTTTGTGTC AAGAGCTAAA ACTGTGAACC TAACTTTCTC	3360
	TTATTGGTGG GTAATAACTG AAAATAAAGA TTTATTTTCA TGCTCACTTC TTAAAAGTCA	3420
	TAAAAACAAT CAAATAGGRT CATGTTTATT GTCATGTGTT TCCTGGKTTT TGACCTGTGT	3480
25	GCACACCCCT GTGTGTTTAT AATTTTAA TGAATTTA TATGGGGTTT TTATTTGCTA	3540
	AAAACCAGGC TGTGAATCA CATTTGGGAA GGGTACTTAT CTTAATGACT AATGACTTAA	3600
30	TGCGGAAAGT TGAATTCTTG TAAAATACAA AATCCAAGGA CTTCTTGGGA TTAAATCTAA	3660
	TTGTCACTTC NTTAGGCAGA TNCAC TTTT TGGATAATGG AAAGTTAAGC ATACCGAATG	3720
	CTACTTTTGG TTGACAAACG GGCCTAATAG TCCGGGGGGA AATCCCTAAC NGGTAAGGNT	3780
35	CCCAAGTATG G	3791

40 (2) INFORMATION FOR SEQ ID NO: 215:

(i) SEQUENCE CHARACTERISTICS:

- 45 (A) LENGTH: 1334 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 215:

50	CAGTGCTCGC TCCTGCTCGG GCGCTGCGG CCCC GGCGT CGCCATGACC AGTGAGCTGG	60
	ACATCTTCGT GGGGAACAGA CCC TTATCGA CGAGGACGTG TATCGCCTCT GGCTCGATGG	120
55	TTACTCGGTG ACCGACGGG TGGCCCTGCG GGTCGCTCG GGAATCCTGG AGCAGACTGG	180
	CGCCACGGCA GCGGTGCTNC AGAGCGACAC CATGGACCAT TACCGCACCT TCCACATGCT	240
	CGAGCGGCTG CTGCATGCGC CGCCAAGCT ACTGCACCAG YTCATCTTCC AGATTCCGCC	300
60	CTCCCGGCAG GCACTACTCA TCGAGAGSTA CTATGCCTTT RATGAGGCCT TTGTTCCGGA	360



	GGTGCTGGGC AAGAAGCTGT CCAAAGGCAC CAAGAAAGAC CTGGATGACA TCAGCACCAA	420
5	AACAGGCATC ACCCTCAAGA GCTGCCGGAG ACAGTTTGAC AACTTTAAAC GGGTCTTCAA	480
	GGTGGTAGAG GAAATGCCGG GCTCCCTGGT GGACAATATT CAGCAACACT TCCTCCTCTC	540
	TGACCGGTTG GCCAGGGACT ATGCAGCCAT CGTCTTCTTT GCTAACAACC GCTTTGAGAC	600
10	AGGGAAGAAA AAAGTGCAGT ATCTGAGCTT CGGTGACTTT GCCTTCTGCG CTGAGCTCAT	660
	GATCCAAAAC TGGACCCCTG GAGCCGTCGA CTCACAGATG GATGACATGG ACATGGACTT	720
15	AGACAAGGAA TTTCTCCAGG ACTTGAAGGA GCTCAAGGTG CTAGTGGCTG ACAAGGACCT	780
	TCTGGACCTG CACAAGAGCC TGGTGTGCAC TGCTCTCCGG GGAAAGCTGG GCGTCTTCTC	840
	TGAGATGGAA GCCAACTTCA AGAACCTGTC CCGGGGGCTG GTGAACGTGG CCGCCAAGCT	900
20	GACCCACAAT AAAGATGTCA GAGACCTGTT TGTGGACCTC GTGGAGAAGT TTGTGGAACC	960
	CTGCCGCTCC GACCACTGGC CACTCAGCGA CGTGCGGTTT TTCTGAATC AGTATTCAGC	1020
25	GTCTGTCCAC TCCTTCGATG GCTTCCGACA CCAGGCCTCT GGGACCGCTA CATGGGCACC	1080
	CTCCGCGGCT GCCTCCTGCG CCTGTATCAT GACTGAGGTG CCTCCCAACG CTCCGCCAC	1140
	GCTGACAATA AAGTTGCTCT GAGTTTGAG ACTGGTCCTC GCTCCGGGGA GCAAGTGGGG	1200
30	GGCGTGCAGA TGTCCTGTG TCTGTCTCTG AGCACCTGGT GTCCGTGTAC AAGGATGGAT	1260
	GTGTNCNGTG GCTCCTTGGG AACTGAGACA TATCTCAGGG AATGGTGTCT GTGCTCAGCC	1320
35	CATCCACCAG AAGA	1334

40 (2) INFORMATION FOR SEQ ID NO: 216:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1511 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 216:

50	GTGGCGGGGA TGCTGCGAGG GGGTCTCCTG CCCAGGCGG GCCGGCTGCC TACCTCCAG	60
	ACTGTCCGCT ATGGCTCCAA GGCTGTTACC CGCCACGTC GTGTGATGCA CTTTCAGCGG	120
	CAGAAGCTGA TGGCTGTGAC TGAATATATC CCCCCGAAAC CAGCCATCCA CCCATCATGC	180
55	CTGCCATCTC CTCCAGCCCC CCCACAGGAG GAGATAGGCC TCATCAGGCT TCTCCGCGG	240
	GAGATAGCAG CAGTTTTCCTA GGACAACCGA ATGATAGCCG TCTGCCAGAA TGTGGCTCTG	300
60	AGTGCAGAGG ACAAGCTTCT TATGCGACAC CAGCTGCGGA AACACAAGAT CCTGATGAAG	360

	RTCTTCCCCA ACCAGGTCTT GAAGCCCTTC CTGGAGGATT CCAAGTACCA AAATCTGCTG	420
	CCCCTTTTTG TGGGGCACAA CATGCTGCTG GTCAGTGAAG AGCCCAAGGT CAAGGAGATG	480
5	GTACGGATCT TAAGGACTGT GCCATTCTTG CCGCTGCTAG GTGGCTGCAT TGATGACACC	540
	ATCCTCAGCA GGCAGGGCTT TATCAACTAC TCCAAGCTCC CCAGCCTGCC CCTGGTGCAG	600
10	GGGAGCTTG TAGGAGGCCT CACCTGCCTC ACAGCCCAGA CCCACTCCCT GCTCCAGCAC	660
	CAGCCCTCC AGCTGACCAC CCTGTGGAC CAGTACATCA GAGAGCAACG CGAGAAGGAT	720
	TCTGTCATGT CGGCCAATGG GAAGCCAGAT CCGTACACTG TTCCGGACTC GTAGCCAGCC	780
15	TGTTTAGCCA GCCCTGCGCA TAAATACACT CTGCGTTATT GGCTGTGCTC TCCTCAATGG	840
	GACATGTGGA AGAACTTGGG GTCGGGGAGT GTGTTTGTC CTTGGTTTTC ACTAGTAATG	900
20	ATATTGTCAG GTATAGGGCC ACTTGGAGAT GCAGAGGATT CCATTTCAGA TGTCAGTCAC	960
	CGGCTTCGTC CTTAGTTTTT CCAACTTGGG ACGTGATAGG AGCAAAGTCT CTCCATTCTC	1020
	CAGGTCCAAG GCAGAGATCC TGAAAAGATA GGGCTATTGT CCCCTGCCTC CTGGTCACT	1080
25	GCCTCTGCT GCACGGGCTC CTGAGCCACC CCCTTGGGGC ACAACCTGCC ACTGCCACAG	1140
	TAGCTCAACC AAGCAGTTGT GCTGAGAATG GCACCTGGTG AGAGCCTGCT GTGTGCCAGG	1200
30	CTTTGTGCTG AGTGCTGTAC ATGTATTAGT TCCTTTACTG CTGACCACAT TGTACCCATT	1260
	TCACAGAGAA GGAGCAGAGA AATTAAAGTG CTGTCTCAAG GTCATGCAGT TAGTAAGTGG	1320
	CAGAACAGGG ACTTGAACCA AGCCCTCTGC TCTGAAGACC GCGTCTGAA TTTCTTCACT	1380
35	AGAGCTTCCT CATCAGGTTA CCCAGAAGTG GGTCCCATCC ACCATCCAGG TGTGCTTGGA	1440
	TGTTAGTTCT CCACCCTCGA GGTGTACGCT GTGAAAAGTT TGGGAGCACT GCTTTATAAT	1500
40	AAAATGAAAT A	1511

## (2) INFORMATION FOR SEQ ID NO: 217:

45

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 642 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

50

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 217:

55

AGGCCTTACT TTTCCTCCCA CAAAGGAGTC GCAGCCACGC TAGCTCTGAC TTGCCACTGT	60
GACAAAGTTC ACGTAGCAGG TCTAGGCAAA GACTGGGCAA TTGAGCAGAG GAGACGGACC	120
TGTGAGTCTG ACCRYGAGSC GGRCCCTTC ACCTTGGCTG GGCTGGTCCT GGTCTTAGG	180
TTTTGTCAGG TTGTCTTGT TTGGATCCCT CAACTAGGTG ATAAGCACTG GAGGGGGATG	240

60

5 ACCCGCCTTG GACGTGTTTC TTAACTCA TCCATATAAT AGGGCCGTGG GATGGTTGTA 300  
GAGGTAAAGC AGGATGATGG TGTTTAAGA CCAGAGCTTG GGACCAGGGC TCCTACACCT 360  
AATTTTCTCT CCTGGTAGCT GAACAAAGGT CTAAATTAGC TTAACAAAAG AACAGGCTGC 420  
CGTCAGCCAG AGTTCTGAAG GCCATGCTTT CAGTTTCCCT TGTGACAAT TGCTCTCCAG 480  
10 TTCCTATGAA AGCACAGAGC CTTAGGGGGC CTGGCCACAG AACACAACCA TCTTAGGCCT 540  
GAGCTGTGAA CAGCAGGGGG TTGTGTGTCT GTTCTGTTTC TCTGCTTGCC GAACTTTCTC 600  
AATAAACCTT ATTTCTTATT TTATATTAC GTGGTGCTG GG 642  
15

20 (2) INFORMATION FOR SEQ ID NO: 218:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1241 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
25 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 218:

30 GGTCCCACTG TTCCATTTTA TGCTAATAGA TTCCATTCTA GGGCCAGCC GTCTCTTGAC 60  
TGATGGTGT CCCTTTAACC CTGGCATGT ATAATAGAAT TTGGTGAAT GAAAGAACCC 120  
AAATAGGCCA GATAGTCCCC CCAGGCCTG ATATCCATAA AAGGCTTGGG AATGCATTAT 180  
35 GTAATTGTCC TTAGTCTTTT TGTGTTTTA GAAAAAACA ACAAGATGGG CTCAGATGGA 240  
TGCCTACGTA AAAATGGTTC CTAGCTGTGT ACTCATAACT TTCTTTGAA TTGAGTAGTG 300  
AAAGGAAGGA GGAGGAAAGG AAATTAAATG TCCTTCTAGT ATTCTCTGGA CTCAAGTCTG 360  
40 ACATATGRGA TAATAACCTA TATTGAAATG CCAAGAATTG TATCTGAAAC AAGGAACAG 420  
TTTGACACAT TTATCATGCC TTCATATTAC ATATTAACCTG AAACCAATTA ATAAACATAT 480  
45 GAAATATCCA TGCACAAGG CAAAGGCACC TAAACCTTTT GTTCTTTTTT CTACATAGCA 540  
GAAATTGATT TTTTTTTTAT TTTTTTAGGG GAACCTATAT AATTATGACC CAGTGATGTC 600  
TTTTGGTGAC TTAAGCTTAT GAATTCAGGT TACAATTGAG TTGATTCTAG ATGTTACTA 660  
50 CCTTGAAAAG GATGTGGTG CCTTATGTGA CACGAGCCAG AGCCTGCTGG GAATAAACAA 720  
AGCAGATTCA TGCCAACACC AACTCGTAGC TTTAGTGGCA GATGGGAGTG GTCACAGACT 780  
55 CCCAAAATGT GGGGCTTTGG ATTTCCACAC CATCCACGT GTGTGTCATC TTCCTCTTTC 840  
AACTCTTGA TGATAATTG AAAATGRTGA AATCACCTCT GAATTGCTT ATAGCATGAG 900  
CACATTCTTA TGACAACATA ACAAATAGTT CATAATGTGA ATATTAGAAA CTGTTACAGC 960  
60

CTGCAGTTAC CATAATTTTC CATGTTTGTG GAATTGATAT TGAAATAGCA GGGCTAAGGA 1020  
ATTACTGGCA AGTTTTAGCC TGTGGGTAAT ACCTTAGGGT TATTTAAATA TTTGTAATTT 1080  
5 TATTTAAATG TTCATGAATG TTTGAAAGGA ACAAATTAT CAGGGATGGC TCTTTGCCAT 1140  
GGGTCTTATT TTCACCTCT TTTCTGTAAG AAAAAAGAAC AATGTCCTAA TGTATTTTAA 1200  
10 AAGTTTTTGG TATAGTTTCT AATTCCAATT TTAATAAAAG T 1241

## (2) INFORMATION FOR SEQ ID NO: 219:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1080 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 219:

25 TGTTTATGTG ACCTAAAACA TACACACATG CACACACACA TACATATCCA TTCATTTCATT 60  
CATTCAAGTG GTGTTTCAG TGTCTGTGTG TCACTGTTTA TGCAGTTTCC ATTTCCCAGT 120  
GAATTATGAG TGGAGGGCAA CTTTCTAAC CAGATTGTCT TTTCAGAACA AAGACCKGGG 180  
30 RATTGAGGAA GAGTTTGAA AGAGGGAGAG GCAAGGAAAG AGAGCTTTAA ATTGAAAGGT 240  
TAATTTCCTA AGAGGAACCT GGGCTGAATG ACTACAGTGT TATACCCTCC AATCTTTGCA 300  
GGTGGGCATG GAACACTGCT TGTATCACTC TGTGCACGGT ATAAATCCAT ATATCCACAA 360  
35 AAACACACAT CCATCCATCA ACATATACAT GGTTTGGGAT GAGCAGGTCA ATAGTTTGA 420  
GAGGGAGTTT GTTCCTTTT TTTCTCATT ATACTCTTAA ATGTTGTCA GTTATCAAAC 480  
40 AAACAAACAG AAAAATGTGTT TGGGAAAAAC CTTCATACG CCTTTTCTAT CMAGTGCTTT 540  
AAAATATAGA CTAAATACAC ACATCCTGCC AGTTTTTCT TACAGTGACA GTATCCTTAC 600  
CTGCCATTTA ATATTAGCCT CGTATTTTTC TCACGTATAT TTACCTGTGA CTTGTATTTG 660  
45 TTATTTAAAC AGGAAAAAAA ACATTCAAAA AAAGAAAAAT TAACTGTAGC GCTTCATTAT 720  
ACTATTATAT TATTATTATT ATTGTGACAT TTGGAATAC TGTGAAGTTT TATCTCTTGC 780  
50 ATATACTTTA TACGGAAGTA TTACGCCTTA AAAATACGAA AATAAATTTT ACAAGGTTTC 840  
TGTTTTGTGT GGAAGAGTAA TTGATGTTGC TAAGAATGAT GTTTGTTTTT TTGGGGTTTT 900  
TGTGTTTTTT TTTTAAATG TTACCAGCAC TTTTTTGTGA AGTTTCACCT TCCGAGGTAT 960  
55 TGTACAAGTT CACACTGTTT GTGAAGTTG AATATGAAGG AATAATTAAA AAAAAAAAAA 1020  
AAACNCGGG GGGGGCCCG TCCCATGGN CCAAGGGG CGGTACGGG GTCACGGCCG 1080  
60

## (2) INFORMATION FOR SEQ ID NO: 220:

5

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1258 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 220:

TGAATTGAGG GCTTAAAGAT AAACATATGG GRTTGGAGTT GTGTGTCCAT AGGGTTTCAC 60

15 TGCCTATTTG ATTTGAGTTT ATCCCTATTA ATTTTTTACA GTGAAATTTT ATTAAAGTAT 120

AATGTACATA TATTTTCAGT GGATTTTGCT CTGAAGGTTT TCCAGTGGTC TGAACACGAG 180

20 ATAGTGGGGC TTCAGCTGTG GGATATTGCA GGGCAGGAGC GCTTCACCTC TATGACACGA 240

TGTATTATC GGGATGCCCTC TGCCCTGTGT ATTATGTTTG ACGTTACCAA TGCCACTACC 300

TTCAGCAACA GCCAGAGGTG GAAACAGGAC CTAGACAGCA AGCTCACACT ACCCAATGGA 360

25 GAGCCGGTGC CCTGCCCTGCT CTTGGCCAAC AAGTGTGATC TGTCCCCTTG GGCAGTGAGC 420

CGGGASCAGA TTGACCGGTT CAGTAAAGAG AACGGTTTCA CAGGTTGGAC AGAAACATCA 480

30 GTCAAGGAGA AAAAAATAT TAATGAGGCT ATGAGAGTCC TCATTGAAAA GATGATGAGA 540

AATTCCACAG AAGATATCAT GTCCTTGTCC ACCCAAGGGG ACTACATCAA TCTACAAACC 600

AAGTCTCCA GCTGGTCTG CTGCTAGTAG TGTGTGGYTT ATTTTCCATC CCAGTTCTGG 660

35 GAGGTCTTTT AAGTCTCTC CCTTTGGTTG CCCACCTGAC MATTTTATTA AGTACATTTG 720

AATTGTCTCC TGAACACTGT CCAGTAAGGA GGCCCATTTG CACTTAGAAA AGACACCTGG 780

40 AACCACAGTG CATTTCTGCA TCTCTGGAT TAGCCTTTSA CATGTTGCTG RCTCACATTA 840

GTGCCAGTTA GTGCCCTGG TGTAAGATCT TCTCATCAGC CCTCAATTTG TGATCCGGAA 900

TTTTGTGAGA AGGATKAGAA ATCAGCACCT GCGTTTTAGA GATCATAATT CTCACCTACT 960

45 TCTGAGCTTA TTTTTCATT TGATATTCAT TGATATCATG ACTTCCAATT GAGAGGAAAA 1020

TGAGATCAAA TGTCATTTCC CAAATTTCTT GTAGGCCGTT GTTTCAGATT CTTTCTGTCT 1080

50 TGAATGTAA ACATCTGATT CTGGAATGCA GAAGGAGGGG TCTGGGCATC TGTGGATTTT 1140

TGGCTACTAG AAGTGTCCCA GAAGTCACTG TATTTTTGAA ACTTCTAACG TCATAATTAA 1200

GTTTCTCTTG TCTTGGGCAT CAAGANTAGT TCCAATTTTT TGGGCCGGGG CAGGGTGG 1258

55

## (2) INFORMATION FOR SEQ ID NO: 221:

60

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1693 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 221:

	CACAATATAT GAAATAGTAC CCTCTAAAAA AGAGAAAAAA AAAATCAGGC GGTCAAACCTT	60
10	AGAGCAACAT TGTCTTATTA AAGCATAGTT TATTTCTACTA GAAAAAATTT AATATCAAGG	120
	ACTATTACAT ACTTCATTAC TAGGAAGTTC TTTTTAAAAAT GACACTTAAA ACAATCACTG	180
15	AAAACTTGAT CCACATCACA CCCTGTTTAT TTTCTTAAA CATCTTGGAA GCCTAAGCTT	240
	CTGAGAATCA TGTGGCAAGT GTGATGGGCA GTAAATACC AGAGAAGATG TTTAGTAGCA	300
	ATTAAAGGCT GTTTGCACCT TTAAGGACCA GCTGGGCTGT AGTGATTCTT GGGGCCAGAG	360
20	TGGCATATG TTTTACAAA ATAATGACAT ATGTCACATG TTTGCATGTT TGTTTGCTTG	420
	TTGAATTTTT GAACAGCCAG TTGACCAATC ATAGAAAGTA TTAATTTCTT TCATATGGTT	480
25	TTTGGTTCAC TGGCTTAAGA GGTTCCTCAG AATATCTATG GCCACAGCAG CATACCAGTT	540
	TCCATCCTAA TAGGAATGAA ATTAATTTTG TATCTACTGA TAACAGAATC TGGGTCACAT	600
	GAAAAAAAT CATTTATCC GTCTTTAAG TATATGTTTA AAATAATAAT TTATGTGTCT	660
30	GCATATGCA GAACAGCTCT GAGAGCAACA GTTCCCATTT AACTCTTTCT GACCAATAGT	720
	GCTGGCACC G TIGCTTCCTC TTTGGGAAGA GGAAAGGGTG TGTGAACATG GCTAACAATC	780
35	TTCAAATACC CAAATTGIGA TAGCATAAAT AAAGTATTTA TTTTATGCCT CAGTATATTA	840
	TTATTTAATT TTTTAGGTAA TGCCTATCTC TTGGTCTATT AAGGAAAGAA GCAATCAGTA	900
	GAGAATTCAG GATAGTTTTG TTTAAATCTT TGCAGATTAC ATGTTTTTAC AGTGGCCTGC	960
40	TATTGAGGAA AGGTATTCTT CYATACAAT TGTTTTAACC TTTGAGAACA TTGACAGAAA	1020
	TTATGCAATG GTTTGTTGAG ATACGGACTT GATGGTGCTG TTTAATCAGT TTGCTTCCAA	1080
45	AGTGGCCTAC TCAAGAGGCC CTAAGACTGG TAGAAATTAA AAGGATTTCA AAAACTTTCT	1140
	ATTCCTTTCT TAAACCTACC AGCAAACCTAG GATTGTGATA GCAATGAATG GTATGATGAA	1200
	GAAAGTTTGA CCAAAATTGT TTTTGTGTG TTGTTGTGT TTTGAATTTG AAATCATTCT	1260
50	TATTCCTTTT AAGAATGTTT ATGTATGAGT GTGAAGATGC TAGCGAACCT ATGCTCAGAT	1320
	ATTCATCGTA AGTCTCCCTT CACCTGTAC AGAGTTTCAG ATCGGTCACT GATAGTATGT	1380
55	ATTTCTTTAG TAAGAATGTG TTAAATTTAC AATGATCTTT TAAAAGATG ATGCAGTTCT	1440
	GTATTTATTG TGCTGTGTCT GGTCTAAGT GGAGCCAATP AAACAAGTTT CATATGTATT	1500
	TTTCCAGTGT TGAATCTCAC AACTGTACT TTGAAAATTT CCTTCCATCC TGAATAACGA	1560
60	ATAGAAGAGG CCATATATAT TGCCTCCTTA TCCTTGAGAT TTCACTACCT TTATGTTAAA	1620

AGTTGTGTAT AATTGTTAAA ATCTGTGAAA GAATAAAAAG TGGATTTAAA TTAAAAAAA 1680

AAAAAAAAA AAA 1693

5

(2) INFORMATION FOR SEQ ID NO: 222:

10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1196 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

15

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 222:

20

ACGCGTGGGT CGACCCACGC GTCCGCGACN TGGCGTGGTG GGAAGGGAG AAGGATTTGT 60

AAACCCCGGA GCGAGGTTCT GCTTACCCGA GGCCGCTGCT GTGCGGAGAC CCCCGGGTGA 120

AGCCACCGTC ATCATGTCTG ACCAGGAGGC AAAACCTTCA ACTGAGGACT TGGGGGATAA 180

25

GAAGGAAGGT GAATATATTA AACTCAAAGT CATTGGACAG GATAGCAGTG AGATTCACCT 240

CAAAGTGAAA ATGACAACAC ATCTCAAGAA ACTCAAAGAA TCATACTGTC AAAGACAGGG 300

30

TGTTCCAATG AATTCACCTCA GGTTCTCTT TGAGGGTCAG AGAATTGCTG ATAATCATAC 360

TCCAAAAGAA CTGGGAATGG AGGAAGAAGA TGTGATTGAA GTTTATCAGG AACAAACGGG 420

GGGTCAATCA ACAGTTTAGA TATCTTTTTC ATTTTTCCTCA ATCCCTTTTTC 480

35

ATTTTAAAA ATAGTCTCTT TGTAAATGTTG TGTTCAAAAC GGAATTGAAA ACTGGCACCC 540

CATCTCTTTG AAACATCTGG TAATTTGAAT TCTAGTGCTC ATTATTCATT ATTGTTTGT 600

40

TTCATTGTGC TGATTTTGG TGATCAAGCC TCAGTCCCCT TCATATTACC CTCTCCTTTT 660

TAAAAATTAC GTGTGCACAG AGAGGTCACC TTTTTCAGGA CATGCAATTT TCAGGCTTGT 720

GGTGATAAAT AAGATCGACC AATGCAAGTG TTCATAATGA CTTTCCAATT GGCCCTGATG 780

45

TTCTAGCATG TGATTACTTC ACTCCTGGAC TGTGACTTTC AGTGGGAGAT GGAAGTTTTT 840

CAGAGAACTG AACTGTGGAA AAATGACCTT TCCTTAACTT GAAGCTACTT TTTAAATTTG 900

50

AGGGTCTGGA CCAAAGAAG AGGAATATCA GGTGGAAGTC AAGATGACAG ATAAGGTGAG 960

AGTAATGACT AACTCCAAG ATGCCTTCAC TGAAGAAAAG GCATTTTAAAG ATTTTAAAA 1020

AATCTGTCA GAAGATCCCA GAAAAGTTCT AATTTTCATT AGCAATTAAT AAAGCTATAC 1080

55

ATGCAGAAAT GAATACAACA GAACACTGCT CTTTTTGATT TTATTGTAC TTTTGGCCT 1140

GGGATATGGG TTTTAAATGG ACATTGTCTG TACCAGCTTC ATTAAATAA ACAATA 1196

60

## (2) INFORMATION FOR SEQ ID NO: 223:

## (i) SEQUENCE CHARACTERISTICS:

5

(A) LENGTH: 1791 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 223:

TCAGGGAGGT	GGCAGGAAAG	GCTTGGAAACA	GCTGCCGGAG	TGACGGAGCG	GCGGCCCCGC	60
CCGTTGCGC	TGGAGGTCGA	AGCTTCCAGG	TAGCGGCCCG	CAGAGCCTGA	CCCAGGCTCT	120
GGACATCCTG	AGCCCCAAGTC	CCCCACACTC	AGTGCACTGA	TGAGTGCGGA	AGTGAAGGTG	180
ACAGGGCAGA	ACCAGGAGCA	ATTCTCTGCTC	CTAGCCAAGT	CGGCCAAGGG	GGCAGCGCTG	240
GCCACACTCA	TCCATCAGGT	GCTGGAGGCC	CCTGGTGTCT	ACGTGTTTGG	AGAACTGCTG	300
GACATGCCCA	ATGTTAGAGA	GCTGGCTGAG	AGTGACTTTG	CCTCTACCTT	CCGGCTGCTC	360
ACAGTGTTTG	CTTATGGGAC	ATACGCTGAC	TACTTAGCTG	AAGCCCGGAA	TCTTCCTCCA	420
CTAACAGAGG	CTCAGAAGAA	TAAGCTTCGA	CACCTCTCAG	TTGTCACCCT	GGCTGCTAAA	480
GTAAAGTGTA	TCCCATATGC	AGTGTGTCTG	GAGGTCTTGC	CCTGCGTAAT	GTGCGGCAGC	540
TGGAAGACCT	TGTGATTGAG	GCTGTGTATG	CTGACGTGCT	TGTTGGCTCC	CTGGACCAGC	600
GCAACCAGCG	GCTCGAGGTT	GACTACAGCA	TGGGGCGGGA	CATCCAGCGC	CAGGACCTCA	660
GTGCCATTGC	CCGAACCCCTG	CAGGAATGGT	GTGTGGGCTG	TRAGGTGCTG	CTGTCAAGCA	720
TTGAGGAGCA	GGTGAGCCCT	GCCAACCAAC	ACAAGGAGCA	GCAGCTGGGC	CTGAAGCAGC	780
AGATTGAGAG	TGAGGTTGCC	AACCTTAAAA	AAACCATTA	AGTTACGACG	GCAGCAGCAG	840
CCGCAGCCAC	ATCTCAGGAC	CCTGAGCAAC	ACCTGACTGA	GCTGAGGGAA	CCAGCTCCTG	900
GCACCAACCA	GCGCCASCCA	GCAAGAAAGC	CTCAAAGGCG	AAGGGGCTCC	GAGGGAGCGC	960
CAAGATTITG	TCCAAGTCGA	ATTGAAAGRA	CTGTGTTTTC	CTCCCTGGGG	ATGTGGGGTC	1020
CCAGCTGCCT	GCTGCTCTCT	TAGGAGTCCT	CAGAGAGCCT	TCTGTGCCCC	TGGCCAGCTG	1080
ATAATCCTAG	GTTTCATGACC	CTTCACCTCC	CCTAACCCTA	AACATAGATC	ACACCTTCTC	1140
TAGGGAGGAG	KCAAATGTAG	GTCATGTTTT	TGTTGGTACT	TTCTGTTTTT	TGTGACTTCA	1200
TGTGTTCCAT	TGCTCCCCGC	TGCCATGCTC	TCTCCCTTGT	TTCTTAAGA	GCTCAGCATC	1260
TGTCCCTGTT	CATTACATGT	CATTGAGTAG	GTGGGTAGCC	CTGATGGGGG	TCGCTCTGTC	1320
TGGAGCATAA	CCCACAGGCG	TTTTTTCTGC	CACCCCATCC	CTGCATGCCT	GATCCCCAGT	1380
TCCTATACCC	TACCCCTGAC	CTATTGAGCA	GCCTCTGAAG	AGCCATAGGG	CCCCCACCTT	1440
TACTCACACC	CTGAGAATTC	TGGGAGCCAG	TCTGCCATGC	CAGGAGTCAC	TGGACATGTT	1500



CATCCTAGAA TCCTGTCACA CTACAGTCAT TTCTTTTCCT CTCTCTGGCC CTTGGGTCCT 1560  
 5 GGGAATGCTG CTGCTTCAAC CCCAGAGCCT AAGAATGGCA GCCGTTTCTT AACATGTTGA 1620  
 GAGATGATTG TTTCTTGGCC CTGGCCATCT CGGGAAGCTT GATGGCAATC CTGGAAGGGT 1680  
 TTAATCTCCT TTTGTGAGTT TGGTGGGGAA GGAAGGGTA TATAGATTGT ATTAAAAAAA 1740  
 10 AAAAGGTATA TATCATATA TCTATATATA ATATGACGCA GAAATAAATC T 1791

15 (2) INFORMATION FOR SEQ ID NO: 224:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2517 base pairs  
 (B) TYPE: nucleic acid  
 20 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 224:

25 ACACTAGTGG ATCCAAAGAA TTCGGCACAG CGGCACAGCA TTGTGAGCT TTTCTGTGTG 60  
 TGTGGGGCCC TCAAGCGAGC TCGACTGGTC CATCCTGGGG TAGCGASGTG GTGTTTGTGA 120  
 AAAAGGACGA TGCCATCACC GCATAYAAGA AGTACAACAA CCGGTGTCTG GACGGGCAGC 180  
 30 CGATGAAGTG CAACCTTCAC ATGAATGGGA ATGTTATCAC CTCAGACCAG CCCATCCTGC 240  
 TCGGCTGAG TGACAGCCCA TCAATGAAAA AGGAGAGCGA GCTGCCTCGC AGGGTGAAC T 300  
 35 CTGCCTCCTC CTCCAACCCC CCTGCCGAAG TGGACCCCTGA CACCATCCTG AAGGCACTCT 360  
 TCAAGTCCTC AGGGGCTCTT KTGACCACGC AGCCACAGA WTTCAAAATC AAGCTTTGAG 420  
 CAGGGGAGTR AGGCAGCCAG AAGTGGGGGC AGAGGAGGGT GGCTCTGTTT CCCCAGGCA 480  
 40 AAGCTTATGA CCAATGGGCC ATCGGACTGG AGACCCCTGA TTGTGGGAAG GGTGTCAGG 540  
 GATAAAGAGC TTCTCACTG GATGGGACCC GCCTTTCTGT GTTGTGTTCT GCCCTGTGCT 600  
 45 CTCTCTCTA CGTTAACGTT TCCTGTAGTA TGTTCTTCA TCTCATCGCC AAGGTAGGCT 660  
 TGTGTTTTTM AGTGTGTGCC TCCCGAGCC TCAGCCCCAA GCTGATTTCT TATCTGGAAA 720  
 TGGTACACTG AATTCTCTGG GTGGCTTTCT TGTGGCCCA TGGGATGCAG CGTGGGGGCT 780  
 50 GTCTGAAGGA CCTGCTTTT TCCAGGGGCC GAGGGGCTGC CTTTCTTTG TGTGTATTAA 840  
 GCTTTTCAAA CAATGGAGGG GATGGAGAGC CCTGGTGTCC TGACGGGAGC CAGGTGCGCC 900  
 55 TGAGAGCTGT GCGCTCCTC TGTCTGTGCA GTGGAGGTGC CTGGGTGGG AGCAGGTCTC 960  
 AGGCCTCTTG TCCTCTCCCC AGTGGCTCCA GGCCTCACTA GTGGCAAGG CAGGATGAGG 1020  
 CTGCACCGCT GGAAGAGTC TATCTAAGCT CTGGCTTGG AGTCCCGTGT CGTCTCCRCC 1080  
 60

CAGAGGAAGT TCTCCAGAGT TCACCTTTCC CTTTTCCTTG AGTTGTGCTG AATGCCCCAC 1140  
 CCCAGCTCTC TTTCCCTTCT GGGTGTCTTT GCTGGGAGGG GGCTGTGTTG TGAGCCCTCC 1200  
 5 CGGTTCAC CTCGCCTGGC ACTTAACCAC ACCCTGGTTT TGTGTAGCCG CCAGCTCTCT 1260  
 TCTGGTTGGG CCTTTGAAAG GCTCAGCCTC CCATTGTGCA GTGCTTGGGT TTGGAGCTTA 1320  
 10 TTTGAATGGA AGAGGTCAGT TTGTTCTGG CTCTCCATT CTGGCCTCAG TTGTCTACAG 1380  
 GACAGTGGTC AGGGATGCCT GGAGGCATAT ATCCAGCTGC CACCAAGGGG CACTGTTTGT 1440  
 TCCCACTTAT GTGAGTGACC CCATCCATCC ATGACCAGAG GATTATTTTC CTGCCTTGGC 1500  
 15 AGAGGAGGAG GAGTCAAGGG AGCAGGGCAG CTCTACCAGG CAAGGTGTTT CCCCAGCATA 1560  
 GGCGCAGACA GTTGGGACGA AACTTCAGAG CCCAGGCAGT CCTGAATGA CCAGGCCAGT 1620  
 GTTGTCACTG AGTGGTCCCC TGCTGGTTGG GAGTGAAGAG AATCCAGGCT GGCAGAGCTG 1680  
 20 GAGCCAGTTG GGGAGCACGG TTCTGGGAGC TCTGCAAAAT CAGTAGCAAG TGCTGAAAAA 1740  
 GGCACATGCC GAAGATACTC AAGAGCTCCC AAGATTTGCT TGAGGCTAGC CCAGTGAAAA 1800  
 25 AAACCAGAGA CTCATGTTTC CAGGGGTCAG TCTGTCAGGC AGGAAGGACC CAGGATTTGA 1860  
 ACCCAGCTTC AGTGTGCAGG CTCTGAGGCT GCCCAGGACG GGAAAGTCCA AGGAAGGGGC 1920  
 CTGGTGGTGC TCCACTTGCA GTTCTTTAAA GAATGCTGCT TTTTATTCTC CTAACCCITT 1980  
 30 CAAGTGGGTG CAGACTTCTC GTTAGCAGCT GGAAGACATT CCTCCACAC TTTTCCCTTC 2040  
 CTGGCCCAAG AGAGCATCCA GAAGGCAGTA GGACCTGGTT TTTCAAGTAC TGGGAGCCGG 2100  
 35 GGGCTCACTG CTTGCACTGT GCTTAGGGTA GGGATGGTAA ATATCCTCCC TGCATGGCTT 2160  
 TATCCTCCCT CTCATCCCAA AGCAGGTATC TTCTGGTTGT CACAGAGTTT CATTGAGTCC 2220  
 AGCTGCAGCC ACGTGGCCAT CTGGAGCTGG TGCTATAGGT GACCATCTGG TACATTGAGG 2280  
 40 GGACCTGTTT GCCTCTCCA CTCTATAAGC AGTCATCTTG GGAGACCGGG AGGAGAAGGT 2340  
 GGTGGGCTAG TCCTGTGCC TCCTCCACTT CCCATGCCTC TATGTTACCC ATCTGTGTCT 2400  
 45 CCTGTGCAGA AGGAGAGGAA GGGGCATTAA GAGATGAAGG GTGATTATGT ATTACTTATC 2460  
 CATTTCTGAA TAAACATTG TTATTCCTAA AAAAAAAAAA AAAAAACTCG AGGGGGG 2517

50

(2) INFORMATION FOR SEQ ID NO: 225:

55

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2424 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 225:

	TTGTANCTAA TOGAGGATTG ATTCTAATGA CAGAGTCTTT CAACACTTTG CACATGATGT	60
5	ATCACGAAGC TACAGCTTGC CATGTGACTG GAGATTTAGT AGAACTTCTG TCAATATTTT	120
	TTTCGGTTTT GAAGTCTACA CGCCCTTATC TTCAGAGAAA AGATGTGAAA CAAGCATTAA	180
	TCCAGTGCCA GGAGCGAATT GAATTTGCCC ATAAACTGTT AACTCTTCTT AATTCCCTATA	240
10	GTCTCCAGA ACTTAGAAAT GCCTGTATAG ATGTCCTCAA GGAAC TTGTA CTTTGTGAGTC	300
	CCCATGATTT TYTTCATACT CTGGTCCCT TTCTACAACA CAACCATTGT ACTTACCATC	360
15	ACAGTAATAT ACCAATGTCT CTGGACCTT ATTTCCCTTG TCRAGAAAAT ATCAAGCTAA	420
	TAGGAGGGAA AAGCAATATT CGGCCTCCGC GCCCTGAACT CAATATGTGC CTCTTGCCCA	480
	CAATGGTGA AACCAGTAAG GGCAAAGATG ACGTTTATGA TCGTATGCTG CTAGACTACT	540
20	TCTTTCTTA TCATCAGTTC ATCCATCTAT TATGCCGAGT TGCAATCAAC TGTGAAAAAT	600
	TTACTGAAAC ATTAGTTAAG CTGAGTGTC TAGTTGCCTA TGAAGGTTTG CCAC TTCATC	660
25	TTGCACTGTT CCCCAACTT TGGACTGAGC TATGCCAGAC TCAGTCTGCT ATGTCAAAAA	720
	ACTGCATCAA GCTTTTGTGT GAAGATCCTG TTTTCGAGA ATATATTAAA TGTATCCTAA	780
	TGGATGAAAG AACTTTT TTA AACAACAACA TTGTCTACAC GTTCATGACA CATTTCTTC	840
30	TAAAGGTTCA AAGTCAAGTG TTTTCTGAAG CAAACTGTGC CAATTTGATC AGCACTCTTA	900
	TTACAAACTT GATAAGCCAG TATCAGAACC TACAGTCTGA TTTCTCCAAC CGAGTTGAAA	960
35	TTTCCAAAGC AAGTGCTTCT TTAATGGGG ACCTGAGGGC ACTCGCTTTG CTCCTGTCAG	1020
	TACACACTCC CAAACAGTTA AACCAGCTC TAATCCAAC TCTGCAAGAG CTTTAAAGCA	1080
	AATGCAGGAC TTGTCTGCAA CAGAGAACT CACTCCAAGA GCAAGAAGCC AAAGAAAGAA	1140
40	AAACTAAAGA TGATGAAGGA GCAACTCCCA TTAAAGGCG GCGTGTAGC AGTGATGAGG	1200
	AGCACACTGT AGACAGCTGC ATCAGTGACA TGAAACAGA AACCAGGAG GTCCTGACCC	1260
45	CAACGAGCAC TTCTGACAAT GAGACCAGAG ACTCCTCAAT TATTGATCCA GGAAGTGAAG	1320
	AAGATCTTCC TTCCCTGAA AATAGTTCTG TTAAAGAATA CCGAATGGAA GTTCCATCTT	1380
	CGTTTTCAGA AGACATGTCA AATATCAGGT CACAGCATGC AGAAGAACAG TCCAACAATG	1440
50	GATAGATAGA CGATTGTAAA GAATTTAAAG ACCTCCACTG TTCCAAGGAT TCTACCCTAG	1500
	CCGAGGAAGA ATCTGAGTTC CCTTCTACTT CTATCTCTGC AGTTCTGTCT GACTTAGCTG	1560
55	ACTTGAGAAG CTGTGATGGC CAAGCTTTGC CCTCCAGGA CCCTGAGGTT GCTTTATCTC	1620
	TCAGTTGTGG CCATTCCAGA GGACTCTTTA GTCATATGCA GCAACATGAC ATTTTAGATA	1680
	CCCTGTGTAG GACCATTGAA TCTACAATCC ATGTCCTCAC AAGGGATATC TGGCAAAGGA	1740
60	AACCAAGCTG CTCTTGACA TTAGGTGTAG CATGTCTACT TTTAAGTCCC TCACCCCAA	1800

CCCCCATGCT GTTTGTATAA GTTTTGCTTA TTTGTTTTTG TGCTTCAGTT TGTCCAGTGC 1860  
 5 TCTCTGCTTG AATGGCAAGA TAGATTTATA GGCTTAATTC TTGGTCAGGC AGAACTCCAG 1920  
 ATGAAAAAAA CTTGCATCTT CAGTATACTT CCTAAAGGC AATCAGATAA TGGATATGTT 1980  
 TTATGTAATT AAGAGTTCAC TTTAGTGGCT TTCATTTAAT ATGGCTGTCT GGAAGAACA 2040  
 10 GGGTTCCTA GCCCTGTACA ATGTAATTTA AACTTACAGC ATTTTACTG TGTATGATAT 2100  
 GGTGTCCTCT GTGCCAGTTT TGTACCTTAT AGAGGCAGAT TGCCTCGAT CGCTGTGGTT 2160  
 CTTATTATCA AAATTAAGTT TACTTGTATA CGGAACAACC ACAAGAAATT TGATTCTGTA 2220  
 15 AAGAATCCTC TTTAGCTGTG GCCTGGCAGT ATATAAATGG TGCTTTATTT AACAGAATAC 2280  
 CTGTGGAGGA AATAAGCAC ACTTGATGTA AAAATAATTG TTTTATTTT ATTGACATGA 2340  
 20 CTGATTGATT GCTATTCTGT GCACTNAATT AACTGATTG TGATGACTTA AAAAAAAAAA 2400  
 AAAAAAAAAA AAAAAAAAAA AAAA 2424

25

(2) INFORMATION FOR SEQ ID NO: 226:

30 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1080 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 226:

ATATAGGACG GATAATCTGT TTACATCTCG TTCTTCTCGA TCACTCACA AGCGGTAAC 60  
 TAGGTGACAA GAAAAAAG ATCTTATTCA AAAGAGGTCT TACAGCAACC CAACGTCTCA 120  
 40 TCTTCCATA GTAAAGATGA CGGCGCCTTG AGGTAAGCTA CAGGCAACAC CACTCCGGG 180  
 TTTCTCTGCG GCGCTGGTCC AAGATGGCGG ATGAAGCCAC GCGACGTGTT GTGTCTGAGA 240  
 45 TCCCGGTGCT GAAGACTAAC GCCGGACCCC GAGATCGTGA GTTGTGGGTG CAGCGACTGA 300  
 AGGAGGAATA TCAGTCCCTT ATCCGGTATG TGGAGAACAA CAAGAATGCT GACAACGATT 360  
 GGTTCCGACT GGAGTCCAAC AAGGAAGGAA CTCGGTGGTT TGGAAAATGC TGGTATATCC 420  
 50 ATGACCTCCT GAAATATGAG TTTGACATCG AGTTTGACAT TCCTATCACA TATCCTACTA 480  
 CTGCCCCAGA AATTGCAGTT CCTGAGCTGG ATGGAAAGAC AGCAAAGATG TACAGGGGTG 540  
 55 GCAAAATATG CCTGACGGAT CATTTCAAAC CTTTGTGGGC CAGGAATGTG CCCAAATTG 600  
 GACTAGCTCA TCTCATGGCT CTGGGGCTGG GTCCATGGCT GGCAGTGAA ATCCCTGATC 660  
 60 TGATTCAGAA GGGCGTCATC CAACACAAAG AGAAATGCAA CCAATGAAGA ATCAAGCCAC 720

	TGAGGCAGGG CAGAGGGACC TTTGATAGGC TACGATACTA TTTTCCTGTG CATCACACTT	780
	AACTCATCTA ACTGCTTCCC CGGACACCCT CCACCTCTAG TTGTTACTAA GTAGCTGCAG	840
5	TAGGCATTGC TGGGGAAGAA ACAAACACAC ACCAAACAGT ACTGCTACTT AGTTTCTAAG	900
	GCTGCACAGG GAAGGGAAG ACTGGGCTTT GGACAATCTA GAGGTAATTT ATATCCGCCC	960
10	CCAGGTGGAG CAACATGCCA TTCTGGAGGC ACGGGGGTAA CTGAAAGTGA GTACATATAG	1020
	TCTTTCTGGT TTCTGGAGAT AACCCATCAA TAAAAGCTGC TTCCTCTGGG TAAAAAAAAG	1080
15	(2) INFORMATION FOR SEQ ID NO: 227:	
	(i) SEQUENCE CHARACTERISTICS:	
20	(A) LENGTH: 1336 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
25	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 227:	
	TTGCATTAC AATTACTGGG AGGCAGGCAG GGGCAGTTGC ATGCTGGGGG TGGCTGCATG	60
	GSCTGCCASC TCTCTGGGT TTGAAGGATG CGGTACASCT GCTTCAGCTG AGCAACGATG	120
30	TTATCCTTGA TGTCGGGGT TGAGATCTGC AGCGGACAC TGCCACTATC AAAGGATCGT	180
	GTGAAATCAC CAGAAAACAT CTCGTAGATC ATCCGAGCCA CTA CTGGAAT GACCTGAACC	240
35	AAGATGAGTT TCCTTTCCAA TGGTTTCCCA TCTGGCCATT CTTCCCCAAA GCATAAGTAG	300
	ATCTCAAACG GTGGCTGCTT CTCTATCTGT CCTTTCTGGT GGGCAATGAG ATCGCTAAGG	360
	AATGTTTCCA GACAAAATAG CTTGACCTTC TTTTGTCTCT CAATCAGGTT GGGAGCAACA	420
40	AGTGATGGGG CACATGGCCC AGACCAGTAC ACCTTGCACT GGCACAGYCT GATGGCATAA	480
	ATGGCATGAC CGCTGACCTC CAGGATCAGT CCTCTGTCCA TGACGTCCAG CAGCTTGCTA	540
45	GTGAACAGCT TCTGCTTCTC ATTGGTAATA TGCTCAGGAC CTGGGAATTT GACCTGCTCC	600
	AGNCTGACGG GACCAAAGAG CTCCTCCTGG TCAGGCATGG GACCCAGGTC CCCATAGAAG	660
	AGTCGGCAGC CCTGAGGGTT GCTCACGGTC ATGGTCCTGC CCGTACTCCT TCCCACGGTA	720
50	CTGAAACTTG ATGTCCAGGT CAGTCATTGG GAGAGAGCTG ATCCACAGTT CTGGAGAGCT	780
	ATAGAAGGRC TGTATAGGTG CCTGGGGWAC TTCCATCTCC AGGGGTTTCTAG TTTTGGGCCA	840
55	CACTGCCTCC GGSCTGCAGT TGCCCACT GCAATTGCCC AACTGGCTG GCGCCATGGG	900
	AGAACCATTG ATGTTCAAGG AGGGGAAGGT GTCTTGATG GGAACATGGT GCTGCGACTG	960
	ATCCAGCTCA TCTTCTCAT CTTCTTCATC CACATCATT TCTTCTCAT CCCAGGAGC	1020
60	AGACCCTGTG GATCCTGGGT TAATGATCGA SCCCTGGGGC TGAGGGATGT CACACACTTG	1080

5 ATATATCTTC ACTGGGTTC TGGGCACCTC CCTTGGTGCC ATCCATACAT CCAGGTGAA 1140  
 TTCTCTGCTC TTATTGAGAG CACAGCGCAG CTGGGCCTTC CATTTAGCTG GGTGAGGTC 1200  
 ATCCACCCCT TCCTGGTACT TCCCTGTCTC TACAGCCCAG GCCTTAAAAA TGGTATTTTC 1260  
 CTCTTCTGTG TGAGGGCTAT GCCGGGTGGC ATGTTTCCAG GGAATCTGGA AGCGTTTGA 1320  
 10 GTCCCTGTGT AGCCAG 1336

15 (2) INFORMATION FOR SEQ ID NO: 228:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2043 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 228:

25 TCAGCTGGTC CCTTCCTGT GTCCCTGGGG ACCTGCTGGC GGCTCTTCC TGGAGCCAT 60  
 GACCTCAGAC CCCACCCACA CTCCAGATCG AGACCCCTGC CTCCTCCCCG CAAATGTCCT 120  
 30 CCCGCTGCCT TGCAGCCTGC ACTTTCACACA TGCTCACCCC CAGCACAGTC CCACTGGCCC 180  
 CTCAMCTCCC CTTCCTGAG CTCTTCCCA AGGACTCCTG GTCAGTGCCT GCTGTGCAKT 240  
 CAGAGGCCCA GGTCCAGCA GCCCGSGGG AACGGGTGCT GCCTSTTCCT CCAGTTAGCT 300  
 35 CCAGYTCAGG TCTGAGACCC GTGYTGAGTA AAGGTCTGAG CAMCGACCGT GCCCTCTGCC 360  
 CAGGGCTGGG TCCTGAGCAG CTGGTTTCC TGCAAGGAAG TTGGAGCAAG CAAAGTCCTT 420  
 CTCTGCCCTC AGGGTCAGCT GCCCAGACTG GGGCGGATGC AGAGAGGCAG GTGGGCTGTG 480  
 40 GCTGGACTGG TCCGGAGCTG GCTTCCTTAC CAGAAAAGCC TCAGCCTTCC TCTGGAAGCA 540  
 TCCCCCGTTC TGGGCAAGGG GGAAGGGCTC CTTTAAGGGG TGTGCTTTCC CAGTGGGGAG 600  
 45 CAGTCTGGCC CTGCCCCCTA CTAAAGCCTC TGCTCTCAGC ACTTTCCTCC AAGTCCTTGT 660  
 AACTTGCTTG AAGGTGGTT CTGGCTGCCA GCCAGTCCCT GGACAACTC TCCTGCCCCCT 720  
 TTTAAATTTC ACTCATTTTG TATAAACCCA GCAGGCTGGT GTTTACTTAG CCCTGTAGCT 780  
 50 TTTTTCATT TTTCTTCCG TCTTCTTCT TGAGTTCACG GTTCAATATT GCCTCCTCGC 840  
 CCTGGTGAGG GGAGGTGCTG CTTTCTGCC CCACCTGCCG GCTGGTTCCA GCAGCGCTGG 900  
 55 NGCCAGCTG GGGGGCCGG ATGGGGCTT CTCTCTCTGG GAGGGGTGCA GGTGCCCTCC 960  
 CCAGGCTGGG AGGGTTCCTT CCCTAGCTCC CCATCTGCCC CCGCTGGTGA GAGTTGGCT 1020  
 60 TCTTGGTCTT GGAATCCCT GGCATTGGGA ACAGAGCATT TCCAGCATTT GTTGTGTGTG 1080

	TTTACTCAC CTAACCCTTA GAAATGAAT GTTAGAAGGT GCCTGCCGAG GCGGGACAGA	1140
	GTGTTTGCTC GCGCTGGAGA AGGCTCTGCT CAGCCCTGAG AGTCCCTTCC TGCCCCACCG	1200
5	ATACTGGCAC TTAAAAAGG AAGCTGACCG CACAGTGTC AGACGAATTG GCCCCAGAA	1260
	GATGGGGAGT TCTGTCTGC CTTCTGTGT CTGCGTGACC TCACCCAGCC TAGGAGGGAG	1320
10	GTGCATTGAG GGTAGATTG CCTCTCATTC AAAGTTCTGG GGCTTTGGGY GGAAAACAGC	1380
	CAGCTTTGGC GCTGTTGGG AGACTCCTCC AGACCAGGAA CCCCAGAAG AGACAGAGCC	1440
	TGCCACATCC TCCACGCCA GGCCCTGGG CAGGGTGATT GGACTGAGAA TTGGCCACA	1500
15	ACCAAATTGA TGCTGGCTGG AACCAGAGGC CAGAAAGCCT GGCTTTGTCC CCATGTGGGA	1560
	GCCCTGTCT CAGCCCTCTT GTCCCTTGA GCTCAGTGAA TTCCACCAG GTGCCACAG	1620
20	CTCCTGGACT TCAAATCTA TATATTGAGA GAGTTGGAGA GTATATCAGA GATATTTTG	1680
	GAAAGGAGTT GGTCTATGCA ATGTCAGTTT GGAATCTTCT TGAAAGTTTA ATGTTTITAT	1740
	TAGGAGATTT AAAGAAAATA AAGGTCTACA ATATCTTAG GTTTTTTTT TTTCTGTTT	1800
25	ACCGCACAAA CTGACCACAT GCCATGTCTA TCAGGATGGA GGGGTCCAT GTTCTCTCT	1860
	GTCTTAGGG AGTGATAAG GAGATGGSCG RAGGGGTGTT TTTTCTTTG ACTCCCTCC	1920
30	TTTCTAACAG AATGTTGCCA CCACTGCTTG AGTGGGCTGT GTTGTTCCT CTGTCCAGC	1980
	TTCTGTGTGA GAAAATAACA TTGTAGGGG AACTCAGGCT AGTGTACGG TCTTGGTTTG	2040
	GGG	2043

35

## (2) INFORMATION FOR SEQ ID NO: 229:

40

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 540 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

45

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 229:

	TAAAAAGAAG CGGGAGAATC TGGCGTCGC TCTAGAGATC GATGGGCTAG AGGAGAAGCT	60
50	GTCCCACTGT CGGAGAGACC TGGAGCCGT GAACTCCAGA CTCCACAGCC GGGAGCTGAG	120
	CCCAGAGGCC AGGAGGTCCC TGGAGAAGGA GAAAACAGC CTAATGAACA AAGCCTCCAA	180
55	CTACGAGAAG GAACTGAAGT TTCTCGGCA AGAGAACCGG AAGAACATGC TGCTCTCTGT	240
	GGCCATCTTT ATCTCTCTGA CGCTCGTCTA TGCTACTGG ACCATGTGAG CTTGGCACTT	300
	CCCCACAACC AGCACAGGCT TCCACTTGGC CCCTTGGTCA GGATCAAGCA GGCACCTCAA	360
60	GCCTCAATAG GACCAAGGTG CTGGGGTGT CCCTCCCAA CCTAGTGTTC AAGCATGGCT	420

5 TCCTGGCGGC CCAGGCCTTG CCTCCCTGGC CTGCTGGGGG GTTCCGGGTC TCCAGAAGGA 480  
CATGGTGCTG GTCCCTCCCT TAGCCCAAGG GAGAGGCAWT AAAGACACAA AGCTGGAAAT 540

10 (2) INFORMATION FOR SEQ ID NO: 230:

- 15 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 448 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 230:  
20 AATTGTGAAA TATTAGAATA TTGTTACTAT TTGACCCAAC TCAAAATCTC CATGGGAAAA 60  
TACCTGTGCA TACCACAGT ATTGTTGAAA ATAATCAGAT GCAGTATCAC AGCTGTGTCA 120  
GACTCTAGTA CCAGTTGGGC AATCAAGGCA CAGCTAAAAA TTGAAAACAA AGATCTGGAC 180  
25 AACAAAACAG CCAAAGGTGG GGGTCAAGAA GCTCTGACGT GTACCTAGCT GTAGAATGCT 240  
ATGCACACGT GCCAGGTGTA GTGTGCATAT CCAGGAAAAA CTGCAGAGAG CCCCAGTCTT 300  
CAMCTCTGGT TGACCATGAG CTCTGTGTAA GCAGGAAGTG AAGGCTAAGG CAGATTTAAG 360  
30 CTCTGAAAGC ATTCCACAAC ATACACACAA ATCGTGCAA GCATTAAGGA AATCTTGTTA 420  
CTGCTAAGTG TTGCTGACCC AGGAACAA 448

35

(2) INFORMATION FOR SEQ ID NO: 231:

- 40 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 407 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
45 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 231:  
GTATGCTGCC CCAAACCAAT ATGTGTGGCT GCCTTTWACC TGACTTCTCC AACATGTAGC 60  
50 CCCAAGAGGA GGCCTCTAGA CTRAGGGAGG GGCTGGTGAC CCAGGTGTGG TGGGGCTGCA 120  
TGARACTACC AGAGAGACAG ACATTCTGGA ACTCACCCCTG GGGGATCCAG TGGATCTGCC 180  
TATGGTCTGG TCCACCCAG ACCTGTGAGA TGTTCCTCAT GAGGATGCAC TTGTGCTTCT 240  
55 GCAAGTATTG CTGCAGCTTC ATAGTGACTC CCACCAGCAC CAGCAATACA GYTAGCTACC 300  
TGTGGCCTTG GATCTCAGCC AGCATGGCTG GGAGAGGGAG CARCTGGGCA TGTACCCTAA 360  
60 ATGCTGTTAC CAGGGAAGGA CTCCAGAGT GAAGACAAGT AGGGACT 407



## 5 (2) INFORMATION FOR SEQ ID NO: 232:

## (i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 830 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 232:

15 GTATTTGATT TCAGGCTGCT AAATGGGCTC ATTTAGCATT CATTCCTGA TGTAGACATT 60  
AAAAAAAA CTGAATAGCA TTCTTTCCAG GNTAACTAAT AAAGCAGACA TGCTAAGCCT 120  
ATAAATACAT CAGCACTGCA GCACACGTTT AAGGTTGCCA CGGACAAGGA TCACACAATA 180  
20 GAGAACACTG TAGTTCGGTC TGCTCACAAG ACCCAGAACA TTGATCAGTT TTTGTTGTTG 240  
GTTTATTATT TTCTGTAA AAAATTGTGA AAAGTTGTT TTAGCTAGAT GATAATTTAA 300  
25 TAGCTGCGAG TGCTTTGGAA CTATAAAGAT GTCACACTT AACACACATA CCTTATGTTT 360  
TGTTTGTGTT TGTTTACAC TCAGTATAAA TCAGGAGAAG TTAGCCAACC ATCTAGCATT 420  
TAGAATCCTC TTTTATTATG TCTCTAAGG ATATGGATGT TCCATAACA GCAACAAAAC 480  
30 AGCAACAAA ACATTCATA AATATCACTT GATAGACTGT AAGCACCTGC TTAACTTTGT 540  
GTCCAAATA TTTAGTGT ATATATATAT ATATATACAC ACACACACAC ATATATATTC 600  
35 AACAAATAA GCAAATATA ACATGCATTT CACATTTTGT CTTCCCTGT TACGATTTTA 660  
ATAGCAGAAC TGTATGACAA GTTAGGTGA TCCTAGCATA TGTTAAATTC AAATTAATGT 720  
AAAACAGATT AACACAACA AAGAACTGT CTATTTGAGT GAAGTCATGC TTTCTATTAT 780  
40 AATAACTTGG CTTGGTTAT CCATCAAATG CACACTTATA CTGTTATCTG 830

45

## (2) INFORMATION FOR SEQ ID NO: 233:

## (i) SEQUENCE CHARACTERISTICS:

- 50 (A) LENGTH: 932 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 233:

55 CCAGAAGAAA GACCAATCTA GAATATGGAA CTCTAATCAC TTCTAGTATT TCAACTTCCT 60  
AGCAGAAATG AACTTGCCC TAGACCTAGG GGATAAGCAA TGTTCTTTAT GTAGCCAATG 120  
60 CTACGGAAAC AAAAGAGGTG AAAGAGACCC TTTTATTATA CTTAATGTAC ATATATTGAC 180

TTTTGTAGCA AGAATGCCAG AAATAGCCTT CATTCTTACC CTGCAAAATA ATCCAGATCT 240  
 5 GCTTCTAAA ATGRANTCAG TTTCTAAAGT GAAACATGCA ATATTTATGC TCTGACTGAC 300  
 TCCTGAATTG GARGAGGAAG RACTTCTGTT TACAGAAAAC YGTATTGTTA TATATGTCAG 360  
 GCTGTGTATT GTGACTATCA GCATTCTGGT GCAAATGAAC TTTTCTCCAT CATCGACTGT 420  
 10 GGAAAATTGA TACTTTTAAA GCATATTCTT CTATGAGCAC AGGTCCCTCT AGTGAAACTT 480  
 AATTTGACAA AGGGTGTCTAT ATGCTTTCCT AACCTGAWTT GTATTAACAT TCACAGAGCC 540  
 TACATTTTCT CATTAGGGTT RTGATGCTCA GTATCTTTCC AAGTGCCAGG CAGRGCTTNC 600  
 15 CTTTCTGAT CAAACATACC ATTTTGTGA TTTCACAACT ATAGACAGTC ACTTCTGCAG 660  
 TCCCAATTTA AAAATGCAGA ACTGCTTTAT CCAAGAATGC TGAAAATAC TGTCTATCC 720  
 20 AGGTTTCCTA AACTATAAAA GCAGATTTTG CTTTGTGTTG TTAATCATAG GCATGGCCGA 780  
 GCATTGTGGA TTAGCTGAG GCTTAAAATC AGATGCATGT CTGGTAAGAT GACCACTGTC 840  
 25 TCACTATCAA GAGCCTGCAG AGCCATTTTC CAGACCTGTG ATTGCCCAGA ACACATAGTC 900  
 CCCACGTTTC TAATTGGAG CAAATCTAAA AG 932

30

(2) INFORMATION FOR SEQ ID NO: 234:

(i) SEQUENCE CHARACTERISTICS:

35

- (A) LENGTH: 2786 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 234:

40

TTAGCAGGT GAGCTGTAA AACAGCACAC ATCTCTCATC CCTCTTCCT TTATTCCTCC 60  
 CTGGGTTTCA GAAAGGAAG ATATATGGGG ACCACCTCCC CCTTCTTTGA TCCCAGCATC 120  
 45 TCAGTCCCCC TCCCAACCTT CCATATGGCT CTCAATGGTG CTCACTTGCT TGGAAGCAGG 180  
 CTCCCAATAG GGAGGGGCT GGCCTCTACA GTCTCTTTGA CTGTAAGACA GGGCTCTGTA 240  
 TCAGTGAGAC GATGAGAAAA GTCCAGGCT AATGGCAGAA ATTTGCACCTT TGAACATGTG 300  
 50 TGTPTTGTG TGTGGGAACC TGAGATTCCT TATTTATTAA CAGGAAGTCT GATTTTTTTT 360  
 TTTTGGAGTC TTTGTGCTA TATTTGTGG GGCTGGGAGA GAGAGATTAG ATTATTTTGA 420  
 55 CATGGGATCC CTTCATAAC AGGTACTTTG AAGGCAAGAC ATAGGGTTGA AGAAGCACAA 480  
 CCAGCCTCTG AAATCATAGC TCTCCAGTGG CTTTAAAGA AAGCTGGTCC TCAGCACTAA 540  
 60 CAAAATCACT ACAATAGCCT AGTGCTTTT TGAAGCCTT TTTAGGGAAG AATGTTAGGT 600

	TCATGGTAAC TAGTATGCTC TTTGAGATTT TTACAGTGTT GAAACTTAAG AATTTTGAGA	660
	GGGTGAGGAG GGTGTTCAG AATCTAAATT ACAGATAGAT GATTGTTTCT TGTGAATTTG	720
5	TTTCTTTTCC TTTTTTTTGG TCCCTACCAT TTCCTTACAT TTCCCTTGGG GCCCATCTCT	780
	GGCTCCTTGC TTTTGTTC TTGCTTTGCT TTATCAGTTC ATTCCAGCTC CTGTTAGTG	840
10	AAGGACACTG CTGTTAGTGA AGGAACAAAG TCTATGAGTC CTAATAATTTT AAGTCAAAGA	900
	AAACTGCTCT GTTCCCTT TAGTAACACT TCTGAAGAGG AAAAATTCA ATAGCCAAAG	960
	TTAATAATCC TATATAATAA TTGCTTTGGC TTTCACCTAA AATTCGGGC ATCACAATTT	1020
15	CCTTGGGATA GAGGTGTGT TGGGAATAG ATTGCTTATT GCTGTCACT GGAGAGAAAA	1080
	GGTAGTGT TTGTACAAGG TCATACCGCC AGAAGCCCCA AATCCTATTT TGGCTCATCT	1140
20	TCAGGTAAAG AGTAATTCCT ATCTGTGTG CTCAGAAGC TAGAATCGAA GGCTTACCCT	1200
	ATTCAITGTT TATTGTGAGA AATGCATGAT GGCTCTTGA AAGAATGACG TTTTGCTGGA	1260
	AAAAAAAAA AGAACAGTTT GTGTTTCACA AACATGGCTT ATCAATTTTT TCAAAGAATT	1320
25	CTTTTTCCT AAAAAGAGGA GTAACAAAT GTCATTCTG AAAGAGGCTT ACTTTATACC	1380
	AACTAGTGT AGCATTTGGG ATGCCAGGA ACAGAGAGTG AGACACCTAC AATCACCAGT	1440
30	CTCAATGCG CTATTGTTTC TTTTCAGAGT GTTGCAGATT TGCCATTTCT CCATAATATG	1500
	GGGATAGAAA ATGGAATAAA GATAGAAGG ATGTAGAATA TGCTTTCCTG CCAACATGGT	1560
	TTGGAGTCGA CTTTGGTATA TTGACTAGAT TTGAAAATAC AAGATTGATT AGATGAATCT	1620
35	ACAAAAAGT TGCTCTCTC TCAGGTCCCT TTTACACTTT TTGACTAACT AGCATCTATA	1680
	TTCCACACT AGCTTTTTTG TCACACTTAT CCTTGTCTC CGTAAATTC ATTTGCAGTG	1740
40	GTTAGTCATC AGATATTTTA GCCACCTACA CAAAAGCAA CTGCATTTT AAAATCTTT	1800
	CTGAGATGGG AGAAAATGTA TTCTCCTTTC CTATACCGCT CTCOCAACAA AAAACAAC	1860
	AGTTAGTTCT ACTAATTAGA AACTTGCTGT ACTTTTCTT TTCTTTTAGG GGTCAAGGAC	1920
45	CCTCTTATA GCTACCATT GCCTACAATA AATTATGCA GCAGTTTGA ATACTAAAAT	1980
	ATTTTATA GACTTTATAT TTTTCTTTT GATAAAGGA TGCTGCATAG TAGAGTTGGT	2040
50	GTAATTAAC TATCTCAGC GTTCCCTGC TTTCCCTTCT GCTCCATATG CCTCATGTG	2100
	CTCCAGGGA GCTCTTTTAA TCTTAAAGT CTACATTCA TGCTCTTAGT CAAATTCTGT	2160
	TACCTTTTA ATAACCTTC CCACTGCATA TTCCATCTT GAATTGGTG TTCTAAATTC	2220
55	TGAACTGTA GTTGAGATAC AGCTATTTAA TATTTCTGGG AGATGTGCAT CCTCTTCTT	2280
	KGTGGTGGC CAAGGTGTT TTGCGTAACT GAGACTCCTT GATATGCTTC AGAGAATTTA	2340
60	GGCAACACT GGCCATGGC GTGGGAGTAC TGGGAGTAAA ATAAAAATAT CGAGGTATAG	2400

ACTAGCATCC ACATAGAGCA CTTGAACCTC CTTTGTACCT GTTTGGGGAA AAAGTATAAT 2460  
 GAGTGTACTA CCAATCTAAC TAAGATTATT ATAGTCTGGT TGTTTGAAAT ACCATTTTIT 2520  
 5 TCTCCTTTTG TGTMTTTCCT ACTTTCCAAT GTACTCAAGA AAATTGAACA AATGTAATGG 2580  
 ATCAATTTAA AATATTTTAT TTCTTAAAAG CCTTTTITGC CTGTTGTAAT GTGCAGGACC 2640  
 10 CTTCTCCTTT CATGGGAGAG ACAGGTAGTT ACCTGAATAT AGGTTGAAAA GGTATGTAA 2700  
 AAAGAAATTA TAATAAAGG GATACTTTGC TTTTCAAATC TTTGTTTCT CTATCTCTAG 2760  
 GTAAGGCATA TTAAAAATAA ATATGT 2786

15

(2) INFORMATION FOR SEQ ID NO: 235:

20

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 458 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 235:

GGGTGCAGGA ATTCGGCAG AGAGAATGTT TGATTTTCTT TCCTATTTTA AGGATCTTCT 60  
 30 CTCTGTGTGA TGTGAAAAC TTACCTTAGT GAAGATGTTT TTCAACATGC TGTGTCTCTT 120  
 TACCTGCATA ATCAGAGCTA TGCATCTATT CAAAGTGATG ATCTGTGGGA TAGTTTAAAT 180  
 GAGGTCACAA ACCAAACACT AGATGTAAAG AGAATGATGA AAACCTGGAC CCTGCAGAAA 240  
 35 GGATTTCTTT TAGTGACTGT TCAAAAGAAA GGAAAGGAAC TTTTATATACA ACAAGAGAGA 300  
 TTCTTTTAA ATATGAAGCC TGAAATTCAG CCTTCAGATA CAAGGTACAT GCCCTCTTTC 360  
 40 TTTTCATGCC ATCTCTTTTG CACTCTCAGG TGGAATATT TTAAAGTGT TTATAATCAT 420  
 AAGTTCTTGT GAAACCTAAC AAGATTATCC CTTCTTAA 458

45

(2) INFORMATION FOR SEQ ID NO: 236:

50

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 591 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 236:

AGGATGAAGA GGAAATTATC TCTTGGATTG CTCTCCAGGA AATCCTTCTC TATACTTTAA 60  
 AAGCTCTTGT TCTTTTCTAG GATCCAAATG TGCTGATTGC TGCTAACAGT CAGGGTACAA 120

60

TTAAGGTGCT AGAATTGGTA TGAAGGGTTA ACTCAAGTCA AATTGTACTT GATCCTGCTG 180  
 AAATACATCT GCAGCTGACA ATGAGAGARG AAACAGAAAA TGTCAATGTA TGTCTCTCCC 240  
 5 CAAAGTCATC ATGGGTTTIG GATTGTGTTT GAATATTTTT TCTTTTTTTC TTKTCCCTCC 300  
 TTTATGAGCC TTTGGGACAT TGGGAATACC CAGCCAACTC TCCACCATCA ATGTAACCTC 360  
 10 ATGGACATTG CTGCTCTTGG TGGTGTATC TAATTTTTGT GATAGGGAAA CAAATCTTTT 420  
 TGAATAAAAA TAAATAACWA AACAATAAAA GTTTATTGAG CCACAGTTGA GCTTGAAAG 480  
 TTTTGTCAA ATGCGCAAG AGATAACTCT TTTTANGAAG TAGCATATGT GAACTATAAT 540  
 15 GTAACAGTGA ATAATTTGTA AAGTCGTAT TTCCCAACCT CTTTGGGAAT T 591

20 (2) INFORMATION FOR SEQ ID NO: 237:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1286 base pairs

(B) TYPE: nucleic acid

25 (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 237:

30 TCTTTTAAAG GTACAGCAGG GAAGAACTGG AAATCAGAG AAAGAACTG CCCTTCCATC 60  
 TACAAAAGCT GAGTTTACTT CTCCTCCTTC TTTGTTCAAG ACTGGGCTTC CACCGAGCAG 120  
 GAGATTACCT GGGGCAATTG ATGTTATCGG TCAGACTATA ACTATCAGCC GAGTAGAAGG 180  
 35 CAGGCGACGG GCAAATGAGA ACAGCAACAT ACAGGTCCTT TCTGAAAGAT CTGCTACTGA 240  
 AGTAGACAAC AATTTTAGCA AACCACCTCC GTTTTTCCTT CCAGGAGCTC CTCCCACTCA 300  
 40 CCTTCCACCT CCTCCATTTT TTCCACCTCC TCCGACTGTC AGCACTGCTC CACCTCTGAT 360  
 TCCACCACCG GGTTTTCCTC CTCCACCAGG CGCTCCACCT CCATCTCTTA TACCAACAAT 420  
 AGAAAGTGGA CATTCCTCTG GTTATGATAG TSGTTCTGCA CGTGCAATTC CATATGGCAA 480  
 45 TCGGATGAAG AACGATACAG ATACAGGGAA TATGCAGAAA GAGGTTATGA GCGTCACAGA 540  
 GCAAGTCGAG AAAANGAAGA ACGACATAGA GAAAGACGAC ACAGGGAGAA AGAGGAAACC 600  
 50 AGACATAAGT CTTCCTGAAG TAATAGTAGA CGTCGCCATG AAAGTGAAGA AGGAGATAGT 660  
 CACAGGAGAC ACAAACACAA AAAATCTAAA AGAAGCAAAG AAGGAAAAGA AGCGGGCAGT 720  
 GAGCCTGCCC CTGAACAGGA GAGCACCGAA GCTACACCTG CAGAATAGGC ATGGTTTTGG 780  
 55 CCTTTTGTGT ATATTAGTAC CAGAAGTAGA TACTATAAAT CTGTGTTATT TTCTGGATAA 840  
 TGTTTAAGAA ATTTACCTTA AATCTGTTC TGTGTGTAG TATGAAAAGT TAACTTTTTT 900  
 60 TCCAAAATAA AAGAGTGAAT TTTTCATGTT AAGTTAAAAA TCTTTGTCTT GTACTATTTT 960

5 AAAAATAAAA AGACAGCAAT GACTTTATAT CCAAGAAAGG AATGTGAATG AGTCACTTAA 1020  
 CAGGGAATCT AAAGAGCTGT GTTAGCTGTG TACATACACA GATTATCTGA GAAAAGGTCA 1080  
 AGGGTTCCAC TTGGGCCACA GTTTTMTTGT TAATCAAACA CCACTCTCTT AAGRGGCTGC 1140  
 ATCACAAARG GCAACCAARG GGGCCCTCTT ARGGCTTTGA GGATTAAAAC TAGTCTTTAT 1200  
 10 CCATTACTGC TGTGGACACT CTGGCTTRG TATWTTTAGG GGGGNTCCTT ACCTTTTMTT 1260  
 GGTTTTCNC ACCTTTTGG TTGGGC 1286

15

(2) INFORMATION FOR SEQ ID NO: 238:

20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 734 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 238:

ATGGCAGCGC AGAAGGACCA GCAGAAAGAT GCCGAGGCGG AAGGGCTGAG CGGCACGACC 60  
 30 CTGCTGCCGA AGCTGATTCC CTCGGTGCA GGCCGGGAGT GGCTGGAGCG GCGCCGCGCG 120  
 ACCATCCGGC CTGGAGCAC CTCGTGGAC CAGCAGCGCT TCTCAGGCC CCGCAACCTG 180  
 GGAGAGCTGT GCCAGCGCCT CGTACGCAAC GTGGAGTACT ACCAGAGCAA CTATGTGTTT 240  
 35 GTGTTCCTGG GCCTCATCCT GTACTGTGTG GTGACGTCCC CTATGTTGCT GGTGGCTCTG 300  
 GCTGTCTTTT TCGGCGCTG TTAACATTCT CTATCTGCGC ACCTGGAGT CCAAGCTTGT 360  
 CCTCTTTGGC CGAAAGGTGA GCCCAGCGCA TCATATGCTC TGGCTGGAGG CATCTCCTTC 420  
 40 CCCTTCTTCT GGCTGGCTGG TGGGGGCTCG GCCGTCTTCT GGTGCTGGG AGCCACCCTG 480  
 GTGGTCATCG GCTCCACGC TGCCTTCCAC CAGATTGAGG CTGTGGACGG GGAGGAGCTG 540  
 45 CAGATGGAAC CCGTGTGAGG TGTCTTCTGG GACCTGCCGG CCTCCCGGC CAGCTGCCCC 600  
 ACCCCTGCCC ATGCCGTGCC TGCACGGTCT GCTGCTCGGG CCCACAGCGC CGTCCCATCA 660  
 CAAGCCCGGG GAGGGATCCC GCCTTTGAAA ATAAAGCTGT TATGGGTGTC ATTCAAAAAA 720  
 50 AAAAAA AAAA 734

55

(2) INFORMATION FOR SEQ ID NO: 239:

60 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 809 base pairs  
 (B) TYPE: nucleic acid

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 239:

5  
CGGGGTCTTC AGGGTACCGG GCTGGTTACA GCAGCTCTAC CCCTCACGAC GCARACATGG 60  
CAGCGCAGAA GGACCAGCAG AAAGATGCCG AGGCGGAAGG GCTGAGCGGC ACGACCCTGC 120  
10 TCCGAAGCT GATTCCCTCC GGTGCAGGCC GGGAGTGGCT GGAGCGGCGC CGCGCGACCA 180  
TCCGGCCCTG GAGCACCTTC GTGGACCAGC AGCGCTTCTC ACGGCCCCGC AACCTGGGAG 240  
AGCTGTGCCA GCGCCTCGTA CGCAACGTGG AGTACTACCA GAGCAACTAT GTGTTCGTGT 300  
15 TCCTGGGCCT CATCTGTAC TGTGTGGTGA CGTCCCCTAT GTTGTGGTG GCTCTGGCTG 360  
TCTTTTTCGG CGCCTGTAC ATTCTCTATC TGCGCACCTT GGAGTCCAAG CTTGTGCTCT 420  
20 TTGCCGAGA GTGAGCCCA GCGCATCAGT ATGCTCTGGC TGGAGGCATC TCCTTCCCCT 480  
TCTTCTGGCT GGCTGGTGGG GGCTCGGCCG TCTTCTGGGT GCTGGGAGCC ACCCTGGTGG 540  
TCATCGGCTC CCACGCTGCC TTCCACCAGA TTGAGGCTGT GGACGGGGAG GAGCTGCAGA 600  
25 TGGAACCCGT GTGAGGTGTC TTCTGGGACC TGCGGCCCTC CGGGGCCAGC TGCCCCACCC 660  
CTGCCCATGC CTGTCTGCA CGCTCTGCT GCTCGGGCCC ACAGCGCCGT CCCATCACAA 720  
30 GCCCGGGGAG GGATCCCGCC TTGAAAATA AAGCTGTTAT GGGTGTCAAT CAGGAAAAAA 780  
AAAAAAAAA AAAAAAAAAA AAAAAAAAAA 809

35

(2) INFORMATION FOR SEQ ID NO: 240:

(i) SEQUENCE CHARACTERISTICS:

40 (A) LENGTH: 2201 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 240:

TGACCCACG CGTCCGGCAA CATGGCGGCT GCCGTGGTGC AGCGCCCGGG CTGAGCGACA 60  
GCAAGTGCA GCGGCTCCTA CCCCAGGTGA GGGGTGGCCT CCGCGTGGGA TCGTGCCCTC 120  
50 TTCAGCCCGC TCCTGTCCCC GACATCACGT GTATTCGCA CGTCCCTCC GCGCTGTGTG 180  
TCTACTGAGA CCGGGAGGCG TGACAGGGCC CCGGTCCCTT CTCAGTGGTG CTCTGTGCTT 240  
55 CAGGGCAAGC TCCCCGTCTC CGGGCGCACT TCCCTCGCCT GTGTTCGGTC CATCTCCTT 300  
TCTCCAGCCT CCTCCCTCG CAGGCGGATG AMCCGGACGA CCGGCCAGTG CCTGGCACCC 360  
CGGGGTGCC ARGGTCCAMG GGAACCCGA AGTCCGAGGA GCGCGARGTC CCGAACCAGG 420  
60

	ARGGGCTGCA GCGCATCAMC GGCCCTGTCTC CCGGCCGTTT GGCTCTCATA GTGGCGGTGC	480
	TGTGCTACAT CAATCTCCTG AACTACATGG ACCGCTTCAC CGTGGCTGGC GTCCTTCCCG	540
5	ACATCGAGCA GTTCTTCAAC ATCGGGGACA GTAGCTCTGG GCTCATCCAG ACCGTGTTC	600
	TCTCCAGTTA CATGGTGTGT GCACCTGTGT TTGGCTACCT GGGTGACAGG TACAATCGGA	660
10	AGTATCTCAT GTGGGGGGG ATTGCCTTCT GGTCCCTGGT GACACTGGGG TCATCCTTCA	720
	TCCCCGAGA GCATTTCTGG CTGCTCCTCC TGACCCGGGG CCTGGTGGGG GTCCGGGAGG	780
	CCAGTTATTC CACCATCGCG CCCACTCTCA TTGCCGACCT CTMTGTGGCC GACCAGCGGA	840
15	COGGATGCTC AGCATCTTCT ACTTTGCCAT TCCGGTGGGC AGTGGTCTGG GCTACATTGC	900
	AGGCTCCAAA GTGAAGGATA TGGCTGGAGA CTGGCACTGG GCTCTGAGGG TGACACCGGG	960
20	TCTAGGAGTG GTGGCCGTTT TGCTGCTGTT CCTGGTAGTG CGGGAGCCGC CAAGGGGAGC	1020
	CGTGGAGCCG CACTCAGATT TGCCACCCCT GAACCCACC TCGTGGTGGG CAGATCTGAG	1080
	GGCTCTGGCA AGAAATCCTA GTTTGCTCCT GTCTTCCCTG GGCTTCACTG CTGTGGCCTT	1140
25	TGTCACGGGC TCCTGGCTC TGTGGGCTCC GGCATTCTTG CTGCGTTCCC GCGTGGTCTT	1200
	TGGGGAGACC CCACCCTGCC TTCCCGGAGA CTCTGCTCTT TCCTCTGACA GTCTCATCTT	1260
30	TGGACTCATC ACCTGCCTGA CCGGAGTCTT GGGTGTGGGC CTGGGTGTGG AGATCAGCCG	1320
	CCGGCTCCGC CACTCCAACC CCCGGGCTGA TCCCCGCTC TGTGCCACTG GCCTCCTGGG	1380
	CTCTGCACCC TTCTCTTCC TGTCCCTTGC CTGCGCCCGT GGTAGCATCG TGGCCACTTA	1440
35	TATTTTATC TTCTATGGAG AGACCTCTCT GTCCATGAAC TGGGCCATCG TGGCCGACAT	1500
	TCTGCTGTAC GTGGTGATCC CTACCCGACG CTCCACCGCC GAGGCCCTCC AGATCGTGCT	1560
40	GTCCACCTG CTGGGTGATG CTGGGAGCCC CTACCTCATT GGCTGATCT CTGACCGCCT	1620
	GCGCGGAAC TGGCCCCCTT CCTTCTTGTG CGAGTTCCGG GCTCTGCAGT TCTCGCTCAT	1680
	GCTCTGCGCG TTTGTTGGGG CACTGGGCGG CGCACTTCC TGGGCACCGC CATCTTCATT	1740
45	GAGGCGGACC GCGGCGGGC ACAGCTGCAC GTGCAGGGCC TGCTGCACGA AGCAGGTCC	1800
	ACAGACGACC GGATTGTGGT GCGCCAGCGG GGCCGCTCCA CCCGCGTGCC CGTGGCCAGT	1860
50	GTGCTCATCT GAGARGCTGC CGCTCACCTA CCTGCACATC TGCCACAGCT GGCCCTGGGC	1920
	CCACCCACG AAGGGCCTGG GCCTAACCCC TTGGCTGGC CCAGCTTCCA GAGGGACCCT	1980
	GGCCCGTGTG CCAGCTCCCA GACACTACMT GGGTAGCTCA GGGGAGGAGG TGGGGGTCCA	2040
55	GGAGGGGAT CCTCTCCAC AGGGGCAGCC CCAAGGGCTC GGTGCTATTT GTAACGAAT	2100
	AAAATTGTGA GCCAGACCCC AGGTGCTGCT TCTGCTCTTT CTCTGGGTGG CCTCTGATCT	2160
60	TGCACCCCGT CTTCACCCCA GGGCTCCTGA AGACTGTGGG T	2201



## (2) INFORMATION FOR SEQ ID NO: 241:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1661 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 241:

5	GTCTTCCCG ACATCGAGCA GTTCTTCAAC ATCGGGGACA GTAGCTCTGG GCTCATCCAG	60
10	ACCGTGTTCA TCTCCAGTTA CATGGTGTG GCACCTGTGT TTGGCTACCT GGGTGACAGG	120
15	TACAATCGGA AGTATCTCAT GTCCGGGGGC ATTGCCTTCT GGTCCCTGGT GACACTGGGG	180
20	TCATSCTTCA TCCCGGAGA GCATTTCTGG CTGCTCCTCC TGACCCGGGG CCTGGTGGGG	240
25	GTCCGGGAGG CCAGTATTTC CACCATGCG CCCACTCTCA TTGCCGACCT CTTTGTGGCC	300
30	GACCAGCGGA SCGGATGCTC AGCATCTTCT ACTTTGCCAT TCCGGTGGGC AGTGGTCTGG	360
35	GCTACATTCG AGGCTCCAAA GTGAAGGATA TGGCTGGAGA CTGGCACTGG GCTCTGAGGG	420
40	TGACACCGGG TCTAGGAGTG GTGGCCGTTT TGCTGCTGTT CCTGGTAGTG CGGGAGCCGC	480
45	CAAGGGGAGC CGTGAGCGC CACTCAGATT TGCCACCCCT GAACCCACC TCGTGGTGGG	540
50	CAGATYTGAG GGCTCTGGCA AGAAATCCTA GTTTCGTCTT GTCTTCCCTG GGCTTCACTG	600
55	CTGTGGCCTT TGTACGGGC TCCCTGGCTC TGTGGGCTCC GGCATTCTG CTGCGTTCCC	660
60	GCGTGGTCTT TGGGAGACC CCACCTGCC TTCCCGAGA CTCCTGCTCT TCCTCTGACA	720
65	GTCTCATCTT TGGACTCATC ACCTGCCTGA CCGGAGTCCT GGGTGTGGGC CTGGGTGTGG	780
70	AGATCAGCCG CCGYTCCGC CACTCCAACC CCCGGGCTGA TCCCCTGGTC TGTGCCACTG	840
75	GCCTCTGGG CTCTGCACCC TTCCTCTTCC TGTCCCTTGC CTGCGCCCGT GGTAGCATCG	900
80	TGGCCACTTA TATTTTCATC TTCATTGGAG AGACCTCCT GTCCATGAAC TGGGCCATCG	960
85	TGGCCGACAT TCTGCTGTAC GTGGTGATCC CTACCCGACG CTCCACCGCC GAGGCCTTCC	1020
90	AGATCGTGCT GTCCACCTG CTGGGTGATG CTGGGAGCCC CTACCTCATT GGCCTGATCT	1080
95	CTGACCGCCT GCGCCGAAC TGGCCCCCTT CTTCTTGTG CGAGTTCCGG GCTCTGCAGT	1140
100	TCTCGTCTAT GCTCTGCGG TTGTGTGGGG CACTGGGCGG CGCACTTTCC TGGGCACCGN	1200
105	CATCTTCATT GAGGCCGACC GCCGGCGGGC ACAGCTGCAC GTGCAGGGCC TGCTGCACGA	1260
110	AGCAGGTGCC ACAGACGACC GGATTGTGGT GCCCAGCGG GGCCGCTCCA CCCGCGTGCC	1320
115	CGTGGCCAGT GTGCTCATCT GAGAGGCTGC CGCTCACCTA CCTGCACATC TGCCACAGCT	1380
120	KGCCCTGGGC CCACCCACG AAGGGCCTGG GCCTAACCCC TTGGCCTGGC CCAGCTTCCA	1440

5 GAGGGACCCCT GGGCCGTGTG CCAGCTCCCA GACACTACMT GGGTAGCTCA GGGGAGGAGG 1500  
 TGGGGGTCCA GGAGGGGGAT CCCTCTCCAC AGGGGNCACC CCAAGGGCTC GGTGCTATTT 1560  
 GTAACGGAAT AAAATTGTGA GCCAGACCCC AGGTGCCTGC TCTCGTCTTT CTCTGGGTGG 1620  
 CCTCTGATCT TGCACCCCGT CTTACCCCCA GGGCTCCTGA A 1661

10

(2) INFORMATION FOR SEQ ID NO: 242:

15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1146 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 242:

NGACAGAAAA GCAGAAGATG AGACTCTGTT CATTCACTTT TCCTAGGCCC ATCCTGTGGT 60  
 25 CATCTTTCCC CCTCCCATCA TACCTCCTCC TTCCTGGAGC CTCTGCCGGC TTGGCTGTAA 120  
 TGGTGGCACT TACCTGGATA TTTCAGTGGG AGGATGAAAG GCGAGACTCA CCCTACGCGG 180  
 30 TGGGACAGAT GGGGAGAGGA AAAAGGCAGA GATNGCCAGG AGAGGGGTGC AGGACAAACC 240  
 AGAGAGGTTG GGTGAGGGA AAAGTGTNGG GAGAAAGTGG GGTGCAGGCC CTGCAGGCCG 300  
 GTTTAGCCAG CAGCTGCGGC CTCCC CGGGC CCTTGGCATC CAACTTCGCA GACAGGGTAC 360  
 35 CAGCCTCCTG GTGTGTATCA TAGGATTTGT TCACATAGTG TTATGCATGA TCTTCGTAAG 420  
 GTTAAGAAGC CGTGGTGGTG CACCATGACA TCCAACCCGT ATATATAAAG ATAAATATAT 480  
 ATATATATGT ATGTAAATTA TAGCACTGAG GGCCCTGCTG CCCTGCTGGA CCAAGCAAAA 540  
 40 CTAAGCCTTT TGGTTTGGGT ATTATGTTTC GTTTTGTAT TTGTTTGT TTGTGGCTTG 600  
 TCTTATGTCG TGATAGCACA AGTGCCAGTC GGATTGCTCT GTATTACAGA ATAGTGT TTT 660  
 45 TAATTCATCA ATGTTCTAGT TAATGTCTAC CTCAGCACCT CCTCTTAGCC TAATTTTAGG 720  
 AGGTGCCCCA ATTTGTGTTT TCAATTTTA CTGGTTACTT TTTGTACAA ATCAATCTCT 780  
 TTCTCTCTTT CTCTCTCCC CACCTCTCAC CCTTGCCCTC TCCATCTCCC TCTCCCGCCC 840  
 50 TCCCCTCTC CTCTGCTC CCCGCTCAT TTCTGTCCAC TCCATTCTCT CTCCCCTCTCT 900  
 CCTGCTCCT GCTGCCCCCT CCCAGCCCA CTTSCCGAG TTGTGCTTGC CGCTCCTTAT 960  
 55 CTGTTCTAGT TCCGAAGCAG TTTCACCTGA AGTTGTGCAG TCCTGGTTGC AGCTTTCCGC 1020  
 ATCTGCCTTC GTTTCGTGTA GATTGACGCG TTTCTTTGTA ATTTCAAGTGT TTCTGACAAG 1080  
 60 ATTTAAAAA AAAAAAGGA AAAAAA AAAAAAAC TCGAGGGGGG GCCCGGTACC 1140

CAATTG

1146

5

(2) INFORMATION FOR SEQ ID NO: 243:

(i) SEQUENCE CHARACTERISTICS:

10

(A) LENGTH: 1350 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 243:

15

AACCACGGC TGCTGCGCA GGGCGTGGAG GGCAGAGGGC CGCGGAGGCG CAGTTGCAAA 60

CATGGCTCAG AGCAGAGACG GCGGAAACCC GTTCGCCGAG CCCAGCGAGC TTGACAACCC 120

20

CTTTCAGCCA CCACCAGCCT ATGAGCCTCC AGCCCCTGCC CCATTGCCTC CACCCTCAGC 180

TCCCTCCTTG CAGCCCTCGA GAAAGCTCAG CCCACAGAA CCTAAGAACT ATGGCTCATA 240

25

CAGCACTCAG GCCTCAGCTG CAGCAGCCAC AGCTGAGCTG CTGAAGAAAC AGGAGGAGCT 300

CAACCGGAAG GCAGAGGAGT TGGACCGAAG GAGNCGAGAG CTGCAGCATG CTGCCCTGGG 360

RGGCAGAGCT ACTCGACAGA ACAATGGCC CCCTCTACCT TCTTTTGTGTC CAGTTCAGCC 420

30

CTGCTTTTTC CAGGACATCT CCATGGAGAT CCCCCAAGAA TTTCAGAAGA CTGTATCCAC 480

CATGTACTAC CTCTGGATGT GCAGCAGST GGCTCTTCTC CTGAACCTCC TCGCCTGCCT 540

35

GGCCAGCTTC TGTGTGAAA CCAACAATGG CGCAGGCTTT GGGCTTTCTA TCCTCTGGGT 600

CCTCCTTTTC ACTCCCTGCT CCTTTGTCTG CTGGTACCGC CCCATGTATA AGGCTTTCCG 660

GAGTGACAGT TCAITCAATT TCTTCGTTTT CTCTTTCATT TTCTTCGTCC AGGATGTGCT 720

40

CTTTGTCTC CAGGCCAATG GTATCCAGG TTGGGGATTC AGTGGCTGGA TCTCTGCTCT 780

GGTGGTCCG AAGGCAACAC AGCAGTATCC GTGCTCATGC TGCTGGTCCG CCTGCTCTTC 840

45

ACTGGCAITG CTGTGCTAGG AATGTGCATG CTGAAACGGA TCCACTCCTT ATACCGCCGC 900

ACAGGTGCCA GCTTTTCAGAA GGGCCAGCAA GAAITTTGCTG CTGGTGTCTT CTCCAACCCT 960

GCGGTGCGAA CCGCARCTTG CCAATGCAGC CGCTGGGGCT GCTGAAAATG CCTTCCGGGC 1020

50

CCCGTGACCC CTGACTGGGA TGCCCTGGCC CTGCTACTTG AGGAGCTGA CTTAGCTCCC 1080

GTCCCTAAGG TCTCTGGGAC TTGGAGAGAC ATCACTAACT GATGGCTCCT CCGTAGTGCT 1140

55

CCCAATCCTA TGGCCATGAC TGCTGAACCT GACAGGCGTG TGGGGAGTTC ACTGTGACCT 1200

AGTCCCCCA TCAGGCCACA CTGCTGCCAC CTCTCACAG CCCCAACCCA GCTTCCCTCT 1260

GCTGTGCCAC GGCTGTGCT TCGGTTATTT AAATAAAAAG AAAGTGAAC TGGAAAAAAA 1320

60

AAAAAAAAA AAAAAAAG GGGGNCNC 1350

## 5 (2) INFORMATION FOR SEQ ID NO: 244:

## (i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 1529 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 244:

15 TCCAGAGGC CGGGGGTTC CAGCTCTGCC TGTAGCAGAG CCCTGAGGAG GAGGAGGAAG 60  
AGGATGTGCT GAAATACGTC CGGGAGATCT TTTTCAGCTA GGCATAAAC TGTGCACTGA 120  
ACTGTCTGCC GAGAGCAGCT GGAGGACAGC TGAGCTTCCA CTGGTGCTGC TGGGCCGMCC 180  
20 GCCTGTGGGA ATGGGGCTCT CTGTGCTCCT ACCTTTGTGC CTTCCTGGGC CTGGCAGATT 240  
CACCTCAGGC CAGAAGCCCC TGGACACTCC GGGCCTTGGG GTGCCGTTCT GAGTGTGCGG 300  
25 AAGGCAGGAC TCAAAATGAG ATCCCATTTG ACTCCCTCTG TATGTACTGT GCCCTCTCCT 360  
GGCTCTTGAG GCTCTGGAGT CCCAATTGTC TGTGTTAGTC AGTGACCAGG TTCCAGGGAA 420  
AATRATGTCA TGTGGTGGTC CAACCTACTG GAACCAAAGA GACAGTACTT TGCAAAGAAA 480  
30 AGGATCACTG CCAGGTGCAC TGAATTGCT ACAGTTTAGT CCGCATGATC TCTCCTGAAG 540  
GAGGAAGCCT GTTTCAAAA TAGTTTCCAT CATGAGTCTA TCAATGAGCT CCCACCTCTC 600  
35 CAGCCAGCCT AGAAAGCAAA CGAGCTGCC ACAGTTCTCT GCCCTGTCTG GGAGGTTGAG 660  
GCCACAGTGT ATAGACTGGT AAGCCAGACA GGCCTCCTCC CGCAAGCTGC TACCTTGCTT 720  
TCACCTGTAC CTTGGTCCCC GGGCAGCTAG CTATAAAGCA AGAGGGACAG GAGCCCAGAA 780  
40 GAGACACTGA GGACAAGAGA TCACACCAGA GTACATGTCT CTGCTCTGT TTTCACTGTG 840  
GCTTTGGACA GGAATATATG AATAAATCAC TGCCATACAG GTTTTCCAAT ACACAAGTGC 900  
45 TAGAAAATAC ACACAATTCC CCAATGCGTA AGTTGTGCTA ATGTCTTTCC AAGTTCTGGG 960  
TTGGGAAGTG GAGGGTGGCA GCGTTTGTTT GTGCGCAACC GTCCAGTCCT GTTCACAGCG 1020  
AGGATTTGGA GTCTCCAGG GTCTCATCAT GGGAGTGATT TGTACCGGA CGCCTCTGCC 1080  
50 CTGTCTGGCT TCAGGTCCAG GGAAGCTTTG AAGCAGTCAA GCCTTGCTTT TGTACCCCAT 1140  
GTGTCTGTCT TTTGTTGAGT CACTCAGAGA TCACTCCTGG ACCTCTGGGG TTGGAGTTCC 1200  
55 AGTGATGGCT TATGGCGGCC CACTCACTAT GGTGGGCTGA GTGGAAGCTC CTTAACCATG 1260  
TCCCCAGAGA CACTGAGGTG CTCGCTCTTT TAATGTCCTC GTTTGTTGCC GTAAGTTCTT 1320  
TGCTAGGTTT CATTTTGGCA TTTGGCAAAT CAGCCTGGAA GTCTGGCCCC ATGACAGCAA 1380  
60

TCACTCCCTC CCCACCCTCC TGAAGCTAGA GGAAGATTG CTCAGATCCA TTAATTAAAG 1440  
CAGGAATTGG TGTGACAATG AGCTGCATGG TTTAGGGAGT CTTTGGGAGC CTGGAAGTC 1500  
5 CTGAAGGACA AACAATCTTG TACTAAGAA 1529

10 (2) INFORMATION FOR SEQ ID NO: 245:

(i) SEQUENCE CHARACTERISTICS:

15 (A) LENGTH: 1537 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 245:

20 GTGCGAGGTC CCCGCCAGCC CCCAGCGGCC TTCCCGGCC GGGCGCTCC CAGAGCAAAC 60  
GAGGCCCTG AGAGCTCCAC CTAGTTCACA GGATAAAATC CCACAGCAGA ACTCGGAGTC 120  
AGCAATGGCT AAGCCCCAGG TGGTGTAGC TCCTGTATTA ATGTCTAAGC TGTCTGTGAA 180  
25 TGCCCTGAA TTTTACCCTT CAGGTTATTC TTCCAGTTAC ACAGAATCCT ATGAGGATGG 240  
TTGTGAGGAT TATCCTACTC TATCAGAATA TGTTCAAGAT TTTTGAATC ATCTTACAGA 300  
GCAGCCTGGC AGTTTGAAG CTGAAATTGA ACAGTTTGCA GAGACCCTGA ATGGTGTGT 360  
TACAACAGAT GATGCTTTGC AAGAACTTGT GGAACATC TATCAACAGG CCACATCTAT 420  
CCCAAATTTC TCTTATATGG GAGCTCGCCT GTGTAATTAC CTGTCCATC ATCTGACAAT 480  
35 TAGCCACAG AGTGGCAACT TCCGCCAATT GCTACTTCAA AGATGTCGGA CTGAATATGA 540  
AGTTAAAGAT CAAGCTGCAA AAGGGATGA AGTTACTCGA AAACGATTTC ATGCATTTGT 600  
40 ACTCTTCTG GGAGAACTTT ATCTTAACCT GGAGATCAAG GGAACAAATG GACAGTTAC 660  
AAGAGCAGAT ATTCTTCAGG TTGGTCTTCG AGAATTGCTG AATGCCCTGT TTTCTAATCC 720  
TATGGATGAC AATTTAATTT GTGCAGTAAA ATTGTTAAAG TTGACAGGAT CAGTTTGGGA 780  
45 AGATGCTTGG AAGGAAAAAG GAAAGATGGA TATGGAAGAA ATTATTCAGA GAATTGAAAA 840  
CGTTGTCCTA GATGCAAACT GCAGTAGAGA TGTAACACAG ATGCTCTTGA AGCTTGTAGA 900  
50 ACTCGGTCA AGTAACTGGG GCAGAGTCCA TGCAACTTCA ACATATAGAG AAGCAACACC 960  
AGAAAATGAT CCTAACTACT TTATGAATGA ACCAACATTT TATACATCTG ATGGTGTTC 1020  
TTTCACTGCA GCTGATCCAG ATTACCAAGA GAAATACCAA GAATTACTTG AAAGAGAGGA 1080  
55 CTTTTTTTCCA GATTATGAAG AAAATGGAAC AGATTTATCC GGGCTGGTG ATCCATACTT 1140  
GGATGATATT GATGATGAGA TGGACCCAGA GATAGAAGAA GCTTATGAAA AGTTTGTTT 1200  
60 GGAATCAGAG CGTAAGCGAA AACAGTAAAG TTAAATTTCA GCATATCAGT TTTATAAAGC 1260

AGTTTAGGTA TGGTGATTTA GCAGAACACA AGAGAGCAAG AAAATGTGTC ACATCTATAC 1320  
CAAATTRAGG ATGTTGAGTT ATGTTACTAA TGTATGCAAC TTAAATTTTG TTAAACACTA 1380  
5 TCTGCCAAAA TAAACTTTAT TCCCTATAAC TTAAATGTG TATATATATA TAATAGTTTA 1440  
TTATGTACAG TTAATCTAC TGTMTTGCT GCAATAAAAT CGATTTTGAA ATAAWRAAA 1500  
10 AAAAAAAAAA AAGGGNGGCC GCTCTAGAGG ANCCAAG 1537

15 (2) INFORMATION FOR SEQ ID NO: 246:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 506 base pairs

(B) TYPE: nucleic acid

20 (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 246:

25 TGCAGGATTT GGCCAGGACC CSCCGCGGTG GCGGTGCTA TCGCTTCGCA GAACCTACTC 60  
AGGCAGCCAG CTGAGAAGAG TTGAGGGAAA GTGCTGCTGC TGGGTCTGCA GACCGCATGG 120  
ATAACGTGCA GCCGAAAATA AAACATCGCC CCTCTGCCT CAGTGTGAAA GGCCACGTGA 180  
30 AGATGCTGCG GCTGGATATT ATCAACTCAC TGGTAACAAC AGTATTCATG CTCATCGTAT 240  
CTGTGTTGGC ACTGATACCA GAAACCACAA CATTGACACT TGGTGGAGGG GTGTTTGCAC 300  
35 TTGTGACAGC AGTATGCTGT CTGCCGACG GGGCCCTTAT TTACCGGAAG CTCTGTGTTCA 360  
ATCCAGCGG TCCTTACCAG AAAAGCCTG TGCATGAAAA AAAAGAAGTT TTGTAATTTT 420  
ATATTACTTT TTAGTTTGAT ACTAAGTATT AAACATATTT CTGKATTATT CCAAAAAAAAA 480  
40 AAAAAAAAAA AAAAAAATT TGGTGG 506

45

(2) INFORMATION FOR SEQ ID NO: 247:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1348 base pairs

50 (B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 247:

55 GTCTTCTTTT TCTGTTTTG AGTTGGTGAG TGAGTGAATA GGGTAACATG GGCCTTCAGG 60  
ATGACCCCTT GGAAGTGTG CGAGTTCTTT AAATCTCAGC TGGATCCTG GACCTGGGAG 120  
60 GCCCCTGTGA GGGCCAGCTC TGGAAAAACC TGGGAGTTGA TGCCGGAGGY TGGGAAGAAC 180

	TCTGCTCGAG GGCAGGGTGC CCTGGAACAC TGGTAGTTCT GGGGCTGGGA GGGAGAGGGG	240
5	CTCCGGCTTT CTCTGAAATG AACACTGCTC TTCAGCAGTT CAAGTACTTG TTCTCAAAAC	300
	ATTTTCTAAT TGATTGGTAG GTTTTCATAA GCATTGTTTC TTAAAGGCAT GGAAAGGGAA	360
	GAATGCTCAA GCAAGTCATG TTIGTTTTCA GTGGGATGGG CCCCGCTTCT CACTGCTGGG	420
10	GGCTTCCCTT TGCATGTGGC ACCTTTGTGC AGGGCCACCA GGCAGACTCT TCCCACCTTC	480
	TCCCCTGAA GCACCAAGGG GCTTGAACCG TAATTTGGCT AATCAGAGGC ATTTTTTTTTG	540
15	TCCTAGTATC TTTCACACTT GTCCAACCGT CTTATTTTTT TAAAAGTTCT GTTGCTTGTA	600
	TTAACACGAA ACTAGAGAGA AATAGTTTCT GAAGCCAGTT TATTGTGAAG ATCCCCAAGG	660
	GGAGGTTCCG TAGAGAAAAA TAGTAAGCTG GTTTAGAAAC TGACGAGGGC AAACAGCCAG	720
20	GACGCATTGG AGAGGAATTT GCCAAAGATC TACCCTGAGA TAACGCCTGT CCAGTGTCTT	780
	CACCACGTGA ATAACCAGCG CTCCAAGTG TTTTCTGCT TTGAAAAAAA AAATTCACA	840
25	AGCTTTTAAA GGTGCATTTA AGAATCCATG TGACTTTAGA ATGGAAGTGC CGGCCCTGGC	900
	AACTGTCAGG TGTGCTAGAA GGTTCGATGC CTCTGGAATG CATGTGATAC TCATCTCCAT	960
	TTTGTTCCTT TGATTGCATT TTTGTTCTTT TAGCAGATCT GTCCCTGTGG GTGGTGTCTA	1020
30	AGAAGTCGGA CACCTTGGTT TTTGTGTTAG ATTGAGCTGG GCAGCTGCAA TCAGCTTCTT	1080
	TATATGCAAA TTAGGCACGA CCCATCTGTG GTTCCCTGGT TGGTGGCTAA TGAAGTGAGG	1140
35	GGAGGGAGGG ATGTCAACCC AAAAGTAGGC CCTCCCATG GCTTTGGCCA GGCCAGACAC	1200
	TTACATCGT TTACATGGTT CTGTGTAATT TTAAAGTTTA TGTGTATAAA GCGAAGCTGT	1260
	TTCTGTGAAA CTGTATATTT TGTAATAAA TATATTGCTA CTTTGAGAWR AAAAAAAAAA	1320
40	AAAAACTCGA GGGGGGCCCG GTACCCAA	1348

45 (2) INFORMATION FOR SEQ ID NO: 248:

(i) SEQUENCE CHARACTERISTICS:

- 50 (A) LENGTH: 1766 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 248:

55	GTGCCGAATC GGCAGAGCGG CACGAGCGGC CACGAGAGCA GGCGGAGTAA AGGGACTTGA	60
	GCGAGCCAGT TGCCGGATTA TTCTATTTC CCTCCCTCTC TCCCGCCCCG TATCTCTTTT	120
60	CACCCCTTCTC CCACCCTCGC TCGGTASCA TGGCGGAGCG TCGCGGCCA CTCAGTCCCA	180

	TTCCATCTCC TCGTCGTCTC TCGGAGCCGA GCCGTCCGCG CCCGGCGGGG GCGGGAGCCC	240
	AGGAGCCTGC CCCGCCCTGG GGACGAAGAG CTGCAGCTCC TCCTGTGCGG TGCACGATCT	300
5	GATTTTCTGG AGAGATGTGA AGAAGACTGG GTTTGTCTTT GGCACCACGC TGATCATGCT	360
	GCTTTCCCTG GCAGCTTTCA GTGTATCAG TGTGGTTTCT TACCTCATCC TGGCTCTTCT	420
10	CTCTGTCACC ATCAGCTTCA GGATCTACAA GTCCGTCATC CAAGCTGTAC AGAAGTCAGA	480
	AGAAGGCCAT CCATTCAAAG CCTACCTGGA CGTAGACATT ACTCTGTCTT CAGAAGCTTT	540
	CCATAATTAC ATGAATGCTG CCATGGTGCA CATCAACAGG GCCCTGAAAC TCATTATTGG	600
15	TCTCTTCTG GTAGAAGATC TGGTTGACTC CTTGAAGCTG GCTGTCTTCA TGTGGCTGAT	660
	GACCTATGTT GGTGCTGTTT TTAACGGAAT CACCTTCTA ATTCTTGCTG AACTGCTCAT	720
20	TTTCAGTGTC CCGATTGTCT ATGAGAAGTA CAAGACCCAG ATTGATCACT ATGTTGGCAT	780
	CGCCCGAGAT CAGACCAAGT CAATTGTTGA AAAGATCCAA GCAAACTCC CTGGAATCGC	840
	CAAAAAAAG GCAGAATAAG TACATGGAAA CCAGAAATGC AACAGTTACT AAAACACCAT	900
25	TTAATAGTTA TAACGTCGTT ACTTGTAATA TGAAGGAAAA TACTCAGTGT CAGCTTGAGC	960
	CTGCATTCCA AGCTTTTITT TTAATTGGT GTTTTCTCCC ATCCTTTCCC TTAAACCTC	1020
30	AGTATCAAGC ACAAAAATTG ATGGACTGAT AAAAGAACTA TCTTAGAACT CAGAAGAAGA	1080
	AAGAATCAAA TTCATAGGAT AAGTCAATAC CTTAATGGTG GTAGAGCCTT TACCTGTAGC	1140
	TTGAAAGGGG AAAGATGGA GGTAAGAGAG AAAATGAAAG AACACCTCTG GGTCTTCTG	1200
35	TCCAGTTTTC AGCACTAGTC TTAATCAGCT ATCCATTATA GTTTTGCCT TAAGAAGTCA	1260
	TGATTAACTT ATGAAAAAAT TATTTGGGA CAGGAGTGTG ATACCTTCCT TGGTTTTTTT	1320
40	TTGCAGCCCT CAAATCCTAT CTTCTGCCC CACAATGTGA GCAGCTACCC CTGATACTCC	1380
	TTTTCTTAA TGATTTAACT ATCAACTTGA TAAATAACTT ATAGGTGATA GTGATAATTC	1440
	CTGATTCCAA GAATGCCATC TGATAAAAA GAATAGAAAT GGAAAGTGGG ACTGAGAGGG	1500
45	AGTCAGCAGG CATGCTGCGG TGGCGTCAC TCCCTCTGCC ACTATCCCCA GGAAGGAAA	1560
	RGCTCCGCCA TTTGGGAAAG TGGTTCTTAC GTCAGTGGAC ACCGGTTCTG AGCAATTAGTT	1620
50	TGAGAACTCG TTCCCGAATG TGCTTTCTC CCTCTCCCT GCCCACCTCA AGTTTAATAA	1680
	ATAAGGTTGT ACTTTCTTA CTATAAATA AAAAAAAAAA AACTCGAGGG GGGCCCGGTA	1740
55	CCCAAATCGC CGGATATGAT CGTAAA	1766

(2) INFORMATION FOR SEQ ID NO: 249:

60

(i) SEQUENCE CHARACTERISTICS:



(A) LENGTH: 2664 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 249:

	AGTGTCTCTCG GAGCAGGCCG AGTAAAGGGA CTTGAGCGAG CCAGTTGCCG GATTATTCTA	60
10	TTTCCCTCC CTCTCTCCG CCCGTATCT CTTTCACCC TTCTCCCACC CTCGCTCGCG	120
	TASCATGGCG GAGCGTCGGC GGCCACTCAG TCCCATTCCT TCTCCTCGTC GTCCTTCGGA	180
	GCCGAGCCGT CCGCGCCCGG CGCGGGCGGG AGCCCAGGAG CCTGCCCCGC CCTGGGGACG	240
15	AAGAGCTGCA GCTCCTCCTG TCGGTGCAC GATCTGATTT TCTGGAGAGA TGTGAAGAAG	300
	ACTGGGTTTG TCTTTGGCAC CACGCTGATC ATGCTGCTTT CCTGGCAGC TTTCAGTGTC	360
20	ATCAGTGTGG TTTCTTACCT CATCCTGGCT CTTCTCTCTG TCACCATCAG CTTCAGGATC	420
	TACAAGTCCG TCATCCAAGC TGTACAGAAG TCAGAAGAAG GCCATCCATT CAAAGCCTAC	480
	CTGGACGTAG ACATTACTCT GTCTCAGAA GCTTTCCATA ATTACATGAA TGCTGCCATG	540
25	GTGCACATCA ACAGGGCCCT GAAACTCATT ATTCGTCTCT TTCTGGTAGA AGATCTGGTT	600
	GACTCCTTGA AGCTGGCTGT CTTCATGTGG CTGATGACCT ATGTTGGTGC TGTPTTTAAC	660
30	GGAATCACCC TTCTAATCTT TGCTGAACTG CTCATTTTCA GTGTCCGAT TGTCTATGAG	720
	AAGTACAAGA CCCAGATTGA TCACTATGTT GGCATCGCCC GAGATCAGAC CAAGTCAATT	780
	GTTGAAAAGA TCCAAGCAAA ACTCCCTGGA ATCGCCAAA AAAAGGCAGA ATAAGTACAT	840
35	GGAAACCAGA AATGCAACAG TTAATAAAC ACCATTTAAT AGTTATAACG TCGTTACTTG	900
	TACTATGAAG GAAATACTC AGTGTCAGCT TGAGCCTGCA TTCCAAGCTT TTTTTTAAAT	960
40	TTGGTGTMTT CTCCCATCCT TTCCCTTTAA CCCTCAGTAT CAAGCACAAA AATTGATGGA	1020
	CTGATAAAG AACTATCTTA GAACTCAGAA GAAGAAAGAA TCAAATTCAT AGGATAAGTC	1080
	AATACCTTAA TGGTGGTAGA GCCTTTACCT GTAGCTTGAA AGGGGAAAGA TTGGAGGTAA	1140
45	GAGAGAAAAT GAAAGAACAC CTCTGGGTCC TTCTGTCCAG TTTTCAGCAC TAGTCTTACT	1200
	CAGCTATCCA TTATAGTTTT GCCCTTAAGA AGTCATGATT AACTTATGAA AAAATTATTT	1260
50	GGGACAGGA GTGTGATACC TTCCTTGGTT TTTTMTTGCA GCCCTCAAAT CCTATCTTCC	1320
	TGCCCCACAA TGTGAGCAGC TACCCCTGAT ACTCCTTTTC TTTAATGATT TAACTATCAA	1380
	CTTGATAAAT AACTTATAGG TGATAGTGAT AATTCTGAT TCCAAGAATG CCATCTGATA	1440
55	AAAAAGAATA GAAATGGAAA GTGGGACTGA GAGGGAGTCA GCAGGCATGC TCGGTGGCG	1500
	GTCATCCCT CTGCCACTAT CCCCAGGGAA GGAAARGCTC CGCCATTTGG GAAAGTGGTT	1560
60	TCTACGTCAC TGGACACCGG TTCTGAGCAT TAGTTTGAGA ACTCGTTCCC GAATGTGCTT	1620

	TCCTCCCTCT CCCCTGCCA CCTCAAGTTT AATAAATAAG GTTGTACTTT TCTTACTATA	1680
5	AAATAAATGT CTGTAAGTGC TGTGCACTGC TGTAAACTTG TTAGAGAAAA AAATAACCTG	1740
	CATGTGGGCT CCTCAGTTAT TGAGTTTTTG TGATCCTATC TCAGTCTGGG GGGGAACATT	1800
	CTCAAGAGGT GAAATACAGA AAGCCTTTTT TTCTTGATCT TTCCCGAGA TTCAAATCTC	1860
10	CGATTCCCAT TTGGGGGCAA GTTTTTTTCT TCACCTTCAA TATGAGAATT CAGCGAAGTT	1920
	GAAAGAAAAA TCATCTGTGA GTTCCTTCAG GTTCTCACTC ATAGTCATGA TCCTTCAGAG	1980
15	GGAATATGCA CTGGCGAGTT TAAAGTAAGG GCTATGATAT TTGATGGTCC CAAAGTACGG	2040
	CAGCTGCAAA AAGTAGTGGA AGGAAATTGT CTACGTGTCT TGGAAAAATT AGTTAGGAAT	2100
	TTGGATGGGT AAAAGGTACC CTTGCCTTAC TCCATCTTAT TTTCTTAGCC CCCTTTGAGT	2160
20	GTTTTAACTG GTTTCATGTC CTAGTAGGAA GTGCATTCTC CATCCTCATC CTCTGCCCTC	2220
	CCAGGAAGTC AGTGATGTC TTTTGGGCT TCCCTCCAA AGGACCTTCT GCAGTGAAG	2280
25	TGCCACATCC AGTTCCTTC TTTTGTGCT GCTGTGTTA GATAATTGAA GAGATCTTTG	2340
	TGCCACACAG GATTTTTTTT TTTTAAAGA AAAACCTATA GATGAAAAAT TACTAATGAA	2400
	ACTGTGTGTA CGTGTCTGTG CGTGCAACAT AAAAATACAG TAGCACCTAA GGAGCTTGAA	2460
30	TCTTGGTTCC TGTAAATTT CAAATTGATG TGGTATTAAT AAAAAAAAAA AAAACAMAA	2520
	AAAAAAAAAA AAAAGGGCGG CCGCTCTAGA GGATCCAAGC TTACGTACGC GTGCATGCCA	2580
35	CGTCCATAGC TCTTCTATA GGGGTCCCCC AAATTCATT CANCGGGCCG TCGGTTTAN	2640
	AAAGGTCGTG ANTGGGGGAA ANCC	2664

40

(2) INFORMATION FOR SEQ ID NO: 250:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 865 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 250:

50

CGTGGGAGTG AGGTACCAGA TTCAGCCCAT TTGGCCCCGA CGCCTCTTCT CTCGGAATCC	60
GGGTGCTGCG GATTGAGGTC CCGGTTCTTA ACGGTGGGAT CGGTGTCTCT GGGATGAGAT	120
55 TTGGCGTTTC CTCGGGGCTT TGGTGGGATC GGTGTCTCTA GGATGAGATT TAGGGTTTCC	180
TCGGGGCTTT CGGGATCTTC ACCTAATATC CGGACTGCAA GATGGAGGAA GGCGGGAACC	240
60 TAGGAGGCCT GATTAARATG GTCCATCTAC TGGTCTGTG AGGTGCCTGG GGCATGCAAA	300

	TGTGGGTGAC CTTGCTCTCA GGCTTCCTGC TTTTCCGAAG CCTTCCCCGA CATACCTTCG	360
	GACTAGTGCA GAGCAAATC TTCCCCTTCT ACTTCCACAT CTCCATGGGC TGTGCCTTCA	420
5	TCAACCTCTG CATCTTGGCT TCACAGCATG CTTGGGCTCA GCTCACATTC TGGGAGGCCA	480
	GCCAGCTTTA CCTGCTGTC CTGAGCCTTA CGCTGGCCAC TGTCAACGCC CGCTGGCTGG	540
	AACCCCGCAC CACAGCTGCC ATGTGGGCCC TGCAAACCGT GGAGAAGGAG CGAGGCCTGG	600
10	GTGGGGAGGT ACCAGGCAGC CACCAGGTC CCGATCCCTA CCGCCAGCTG CGAGAGAAGG	660
	ACCCCAAGTA CAGTGCTCTC CGCCAGAATT TCTTCCGCTA CCATGGGCTG TCCTCTCTTT	720
15	GCAATCTGGG CTGCGTCTG AGCAATGGGC TCTGTCTGCG TGGCCTTGCC CTGGAAATAA	780
	GGAGCCTCTA GCATGGGCCC TGCATGCTAA TAAATGCTTC TTCAGAAAAA AAAAAAAAAA	840
20	AAACTCGAGG GGGGCCCGGT ACCCA	865

25 (2) INFORMATION FOR SEQ ID NO: 251:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2082 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 251:

35	TGGGGGGGN AATGGGTGTC TGGCTCANGG ATTGCCNAAT CTGGAAATTC TCCATAACTT	60
	GCTAGCTTGT TTTTITTTTT TTTTITTTACA CCCCCCGCC CCACCCCGG ACTTGACAA	120
	TGTTCAATGA TCTCAGCAGA GTTCTTCATG TGAAACGTTG ATCACCTTTG AAGCCTGCAT	180
40	CATTACATA TTTTITCTTC TTCTTCCCCT TCAGTTCATG AACTGGTGT CATTTTCTGT	240
	GTGTGTGTGT GTTTATTTTT GTTTGGATTT TTTTITTTAA TTTTACTTTT AGAGCTTGCT	300
45	GTGTGCCCCA CCTTTTITCC AACCTCCACC CTCACTCCTT CTCAACCCAT CTCTTCCGAG	360
	ATGAAAGAAA AAAAAAGCA AAGTTTTTTT TTCTTCTCCT GAGTTCTTCA TGTGAGATTG	420
	AGCTTGCAAA GGAATAAAAA ATGTGAAATG TTATAGACTT GCAGCGTGCC GAGTTCCATC	480
50	GGTITTTTTT TTTAGCATTG TTATGCTAAA ATAGAGAAAA AAATGCTCAT GAACCTTCCA	540
	CAATCAAGCC TGCATCAACC TTCTGGGTGT GACTTGTGAG TTTTGGCCTT GTGATGCCAA	600
55	ATCTGAGAGT TTAGTCTGCC ATTAAAAAAA CTCATTCTCA TCTCATGCAT TATTATGCTT	660
	GCTACTTTGT CTTAGCAACA ATGAACTATA ACTGTTTCAA AGACTTTATG GAAAAGAGAC	720
	ATTATATTAA TAAAAAAA AAGCCTGCAT GCTGGACATG TATGGTATAA TTATTTTTTC	780
60	CTTTTTTTTT CCTTTTGGCT TGGAAATGGA CGTTCGAAGA CTTATAGCAT GGCATTCTA	840

	CITTTGTTTT ATGCCTCAT GACTTTTTTG AGTTTAGAAC AAAACAGTGC AACCGTAGAG	900
5	CCTTCTTCCC ATGAAATTTT GCATCTGCTC CAAACTGCT TTGAGTTACT CAGAACTTCA	960
	ACCTCCCAAT GCACTGAAGG CATTCTTGT GCAAAGATAC CAGAATGGGT TACACATTTA	1020
	ACCTGGCAAA CATGAAGAA CTCCTRATGT TTCTTTTTA ATAAGAATGA CGCCCCACTT	1080
10	TGGGGACTAA AATTGTGCTA TTGCCGAGAA GCAGTCTAAA ATTTATTTTT TAAAAAGAGA	1140
	AACTGCCCCA TTATTTTGG TTGTTTTAT TTTATTTA TATTTTGG CTTTGGTCA	1200
15	TTGTCAAATG TGAATGCTC TGGGTTTCTA GTATATAATT TAATTCTAGT TTTTATAATC	1260
	TGTTAGCCCA GTTAAATGT ATGCTACAGA TAAAGGAATG TTATAGATAA ATTTGAAAGA	1320
	GTTAGGTCTG TTTAGCTGTA GATTTTTTAA ACGATTGATG CACTAAATTG TTTACTATTG	1380
20	TGATGTTAAG GGGGGTAGAG TTGCAAGGG GACTGTTTAA AAAAAGTAGC TTATACAGCA	1440
	TGTGCTTGCA ACTTAAATAT AAGTTGGTA TGTGTAGTCT TTGCTATACC ACTGACTGTA	1500
25	TTGAAAACCA AAGTATTAAG AGGGGAAACG CCCCTGTTA TATCTGTAGG GGTATTTTAC	1560
	ATTCAAAAAT GTATGTTTTT TTTCTTTTC AAAATTAAAG TATTTGGGAC TGAATTGCAC	1620
	TAAGATATAA CTGCAAGCA TATAATACAA AAAAAAATG CAAACTGTT TAGAACGCTA	1680
30	ATAAAATTTA TGCAGTTATA AAAATGGCAT TACTGCACAG TTTTAAGATG ATGCAGATTT	1740
	TTTTACAGTT GTATTGTTGT GCAGAACTGG ATTTCTGTGTA ACTTAAAAA AAATCCACAG	1800
35	TTTTAAAGGC AATAATCAGT AAATGTTATT TTCAGGGACT GACATCCTGT CTTTAAAAAG	1860
	AAATGAAAAG TAAATCTTAC CACAATAAAT ATAAAAAAT CTGTTCAGTT ACTTTTCTTT	1920
	TACATATTTT GCTGTGCAAA ATTGTTTTAT ATCTTGAGTT ACTAACTAAC CACGGGTGTT	1980
40	GTTCCTATGT GCTTTCTTT CATTTTCAAT TCTGGTTATA TCAAGAAAAG AATAATCTAC	2040
	AATAATAAAC GGCATTTTTT TTTGAAAAA AAAAAAAAAA AA	2082

45

(2) INFORMATION FOR SEQ ID NO: 252:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1482 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 252:

	CAGGCAGGCT GGCCCCGGG ACTTCTCTCT GGCCCTGCTC CCTCCGAGCG CTCGCCGTT	60
60	GGCCGCCCTGG CCCCTACGGA GTCCTTAGCC AGGATGGAGG CTGTTGTGAA CTTGTACCAA	120

GAGGTGATGA AGCACGCAGA TCCCCGGATC CAGGGCTACC CTCTGATGGG GTCCCCCTTG 180  
CTAATGACCT CCATTCTCCT GACCTACGTG TACTTCGTTT TCTCACTTGG GCCTCGCATC 240  
5 ATGGCTAATC GGAAGCCCTT CCAGCTCCGT GGCTTCATGA TTGTCTACAA CTTCTCACTG 300  
GTGGCACTCT CCTCTACAT TGTCTATGAG TTCCTGATGT CGGGCTGGCT GAGCACCTAT 360  
ACCTGGCGCT GTGACCTGT GGAATATTC AACAGCCCTG AGGCACTTAG GATGGTTCCG 420  
10 GTGGCCTGGC TCTTCTCTT CTCCAAGTTC ATTGAGCTGA TGGACACAGT GATCTTTATT 480  
CTCCGAAAGA AAGACGGGCA GGTGACCTTC CTACATGTCT TCCATCACTC TGTGCTTCCC 540  
15 TGGAGCTGGT GGTGGGGGT AAAGATTGCC CCGGGAGGAA TGGGCTCTTT CCATGCCATG 600  
ATAAACTCTT CCGTGCATGT CATAATGTAC CTGTACTACG GATTATCTGC CTTTGGCCCT 660  
GTGGCACAAC CCTACCTTTG GTGGAAAAAG CACATGACAG CCATTCACTG GATCCAGTTT 720  
20 GTCTGGTCT CACTGCACAT CTCCCAGTAC TACTTTATGT CCAGCTGTAA CTACCAGTAC 780  
CCAGTCATTA TTCACCTCAT CTGGATGTAT GGCACCATCT TCTTCATGCT GTTCTCCAAC 840  
25 TTCTGGTATC ACTCTTATAC CAAGGGCAAG CGGCTGCCCC GTGCACTTCA GCAAAATGGA 900  
GCTCCAGGTA TTGCCAAGGT CAAGGCCAAC TGAGAAGCAT GGCCTAGATA GGCGCCACC 960  
TAAGTGCTC AGGACTGCAC CTTAGGGCAG TGTCCGTCAG TGCCCTCTCC ACCTACACCT 1020  
30 GTGACCAAGG CTTATGTGGT CAGGACTGAG CAGGGGACTG GCCCTCCCTT CCCACAGCT 1080  
GCTCTACAGG GACCACGGCT TTGGTTCTTC ACCCACTTCC CCCGGGCAGC TCCAGGGATG 1140  
35 TGGCCTCATT GCTGTCTGCC ACTCCAGAGC TGGGGGCTAA AAGGGCTGTA CAGTTATTTT 1200  
CCCCCTCCCTG CCTTAAACT TGGGAGAGGA GCACTCAGGG CTGGCCCCAC AAAGGGTCTC 1260  
GTGGCCTTTT TCCTCACACA GAAGAGGTCA GCAATAATGT CACTGTGGAC CCAGTCTCAC 1320  
40 TCCTCCACCC CACACACTGA AGCAGTAGCT TCTGGGCCAA AGGTCAAGGT GGGCGGGGGC 1380  
CTGGGAATAC AGCCTGTGGA GGCTGCTTAC TCAACTTGTG TCTTAATTAA AAGTGACAGA 1440  
45 GGAAACCAAA AAAAAAAAAA AAAAATCGA GGGGGGCCG TA 1482

50 (2) INFORMATION FOR SEQ ID NO: 253:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 834 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 253:

60 GGACAGACCG CCGTTGCCG CTTGGCCCTT ACGGAGTCT TAGCCAGGAT GGAGGCTGTT 60

GTGAACCTGT ACCAAGAGGT GATGAAGCAC GCAGATCCCC GGATCCAGGG CTACCCCTCTG 120  
ATGGGGTCCC CCTTGCTAAT GACCTCCATT CTCCTGACCT ACGTGTACTT CGTTCCTCTCA 180  
5 CTTGGGCCTC GCATCATGGC TAATCGGAAG CCCCTCCAGC TCCGTGGCTT CATGATTGTC 240  
TACAACCTCT CACTGGTGGC ACTCTCCCTC TACATTGTCT ATGAGTTCTT GATGTGCGGC 300  
10 TGGCTGAGCA CCTATACCTG GCGCTGTGAC CCTCAGGACT GCACCTTAGG GCAGTGTCGG 360  
TCAGTGCCTT CTCCAMCTAC ACCTGTGACC AAGGCTTATG TGGTCAGGAC TGAGCAGGGG 420  
ACTGGCCCTC CCCTCCCCAC AGCTGCTCTA CAGGGACCAC GGCTTTGGTT CCTCACCAC 480  
15 TTCCCCGGG CAGCTCCAGG GATGTGGCCT CATGTCTGTC TGCCACTCCA GAGCTGGGGG 540  
CTAAAAGGGC TGTACAGTTA TTTCCCCCTC CTGCTTTAA AACTTGGGAG AGGAGCACTC 600  
20 AGGGCTGGCC CCACAAAGGG TCTCGTGGCC TTTTCTCTCA CACAGAAGAG GTCAGCAATA 660  
ATGTCACTGT GGACCCAGTC TCACTCCTCC ACCCCACACA CTGAAGCAGT AGCTTCTGGG 720  
CCAAAGGTCA GGGTGGGCGG GGGCTGGGA ATACAGCCTG TGGAGGCTGC TTAICTCACT 780  
25 TGTGTCTTAA TTAAAAGTGA CAGAGGAAAC CACGAAAAA AAAAAAAAAA AAAA 834

30

(2) INFORMATION FOR SEQ ID NO: 254:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1508 base pairs  
35 (B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 254:

TTGAACCTTT AAAATTTTAG ATCAGCAAAC TCTAAGATCC TAGAATGGAA GCTGTTCCTC 60  
ATTTCTCCAT GCTCACCTC CCAGGTCAGC GAGATGGTGA AGAAGCTGCA CGCGGCAACA 120  
45 CCACCAACGT TCGGAGTGA CCTCATCAAT GAGCTTGTGG AGAACTTTGG CAGATGTCCC 180  
AAGTGGTCTG GTCGGCAAGC CTTTGTCTTT GTCTGCCAGA CTGTCAITGA GGATGACTGC 240  
CTTCCCATGG ACCAGTTTGC TGTGCATCTC ATGCCGCATC TGCTAACCTT AGCAAATGAC 300  
50 AGGGTTCCTA ACGTGCAGT GCTGCTTGCA AAGACATTAA GACAACTCT ACTAGAAAAA 360  
GACTATTTCT TGGCCTCTGC CAGCTGCCAC CAGGAGGCTG TGGAGCAGAC CATCATGGCT 420  
55 CTTCAGATGG ACCGTGACAG CGATGTCAAG TATTTTGCAA GCATCCACCC TGCCAGTACC 480  
AAAATCTCCG AAGATGCCAT GAGCACAGC TCCTCAACCT ACTAGAAGGC TTGAATCTCG 540  
GTGTCTTTCC TGCTTCCATG AGAGCCGAGG TTCAGTGGC ATCGCCACG CATGTGACCT 600  
60

	GGGATAGCTT TCGGGGGAGG AGAGACCTTC CTCTCCTGCG GACTTCATMG CAGGTGCAAG	660
	TTGCCTACAC CCAATACCAG GGATTTCAG AGTCAAGAGA AAGTACAGTA AACACTATTA	720
5	TCCTATCTTG ACTTTAAGGG GAAATAATTT CTCAGAGGAT TATAATTGTC ACCGAAGCCT	780
	TAAATCCTTC TGTCTTCCTG ACTGAATGAA ACTTGAATTG GCAGAGCATT TTCTTATGG	840
10	AAGGGATGAG ATTCCCAGAG ACCTGCATTG CTTTCTCCTG GTTTTATTTA ACAATCGACA	900
	AATGAAATTC TTACAGCCTG AAGGCAGACG TGTGCCCAGA TGTGAAAGAG ACCTTCAGTA	960
	TCAGCCCTAA CTCTTCTCTC CCAGGAAGGA CTGTCTGGGC TCTGTGCCCA GCTGTCCAGC	1020
15	CCAGCCCTGT GTGTGAATCG TTTGTGACGT GTGCAAATGG GAAAGGAGGG GTTTTACAT	1080
	CTCTAAAGG ACCTGATGCC AACACAAGTA GGATTGACTT AAACCTTTAA GCGCAGCATA	1140
20	TTGCTGTACA CATTTACAGA ATGGTTGCTG AGTGTCTGTG TCTGATTTTT TCATGCTGGT	1200
	CATGACCTGA AGGAAATTTA TTAGACGTAT AATGTATGTC TGGTGTTTTT AACTTGATCA	1260
	TGATCAGCTC TGAGGTGCAA CTTCTTCACA TACTGTACAT ACCTGTGACC ACTCTTGGA	1320
25	GTGTGCACT CTTTAATCAT GCTGTTTAAA CTGTGTGGC ACAAGTCTC TTGTCCAAAT	1380
	AAAATTTATT AATAAGATCT ATAGAGAGAG ATATATACAC TTTTGATTGT TTTCTAGATG	1440
30	TCTACCAATA AATGCAATTT GTGACCTGTA TTAATAAAAA NTAAAAAAC TCGAGGGGGG	1500
	CCCGGTAC	1508

35

(2) INFORMATION FOR SEQ ID NO: 255:

(i) SEQUENCE CHARACTERISTICS:

40

- (A) LENGTH: 2514 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 255:

45

	GAGAGACTCA CACTTCTTTT CCATTATCAC TGACGATGTA GTGGACATAG CAGGGGAAGA	60
	GCACCTACCT GTGTGGTGA GGTGTGTGA TGAATCTCAT AACCTAAGAG AGGAATTTAT	120
50	AGGCTTCCTG CCTTATGAAG CCGATGCAGA AATTGTGGCT GTGAAATTC AACTATGAT	180
	AACTGAGAAG TGGGATTAA ATATGGAGTA TTGTCTGGC CAGGCTTACA TTGWCTCTAG	240
	TGGATTTTCT TCCAAATGA AAGTTGTGC TTCTAGACTT TTAGAGAAAT ATCCCCAAGC	300
55	TATCTACACA CTCTGCTCTT CCTGTGCTT AAATATGTGG TTGGCAAAT CAGTACCTGT	360
	TATGGGAGTA TCTGTGCAT TAGGAACAAT TGAGGAAGTT TGTCTTTTT TCCATCGATC	420
60	ACCACAACCTG CTTTATAGAAC TTGACAACGT AATTCTGTCT CTTTTTCAGA ACAGTAAAGA	480

	AAGGGGTAAA GAACTGAAGG AAATCTGCCA TTCCTAGTGG ACAGGCAGGC ATGATGCTTT	540
5	TGAAATTTTA GTGGAAGTCC TGCAAGCACT TGTTTTATGT TTAGATGGTA TAAATAGTGA	600
	CACAAATATT AGATGGAATA ACTATATAGC TGGCCGAGCA TTGTACTCTC GCAGTGCAGT	660
	GTCAGATTTT GATTTCATG TTAATATTGT TGTCTTAAA AATGTCTAT CTTTTACAAG	720
10	AGCCTTTGGG AAAAACCTCC AGGGGCAAAC CTCTGATGTC TTCTTTGCGG CCGGTAGCTT	780
	GACTGCAGTA CTGCATTCAC TCAACGAAGT GATTGGAATA TATTGAAGTT TATCATGAAT	840
15	TTTGGTTTGA GGAAGCCACA AATTTGGCAA CCAAACTTGA TATTCAAATG AAATCCCTG	900
	GGAAATTCGG CAGAGCTCAC CAGGGTAACT TGAATCTCA GCTAACCTCT GAGAGTTACT	960
	ATAAGAAAC CCTAAGTGC CCAACAGTGG AGCACATTAT TCAGGAAGTT AAAGATATAT	1020
20	TCTCAGAACA GCACCTCAAA GCTCTTAAAT GCTTATCTCT GGTACCCTCA GTCATGGGAC	1080
	AACTCAAATT CAATACGTCG GAGGAACACC ATGCTGACAT GTATAGAAGT GACTTACCCA	1140
25	ATCCTGACAC GCTGTGAGCT GAGCTTCATT GTTGGAGAAT CAAATGGAAA CACAGGGGGA	1200
	AAGATATAGA GCTTCCGTCC ACCATCTATG AAGCCCTCCA CCTGCCTGAC ATCAAGTTTT	1260
	TTCCTAATGT GTATGCATG CTGAAGGTCC TGTGTATTCT TCCTGTGATG AAGGTTGAGA	1320
30	ATGAGCGGTA TGAATATGGA CGAAGCGTC TTAAAGCATA TTTGAGGAAC ACTTTGACAG	1380
	ACCAAAGGTC AAGTAAGTGG GCTTTGCTTA ACATAAATTT TGATATAAAA CACGACCTGG	1440
35	ATTTAATGGT GGACACATAT ATTAACTCT ATACAAGTAA GTCAGAGCTT CCTACAGATA	1500
	ATTCCGAAAC TGTGGAAAAT ACCTAAGAGA CTTTTAAAA TAGGCTTTCT TATATTTGAT	1560
	ATTTGGAAGA AAAAGCCGTA AGTGTATGTA GACCACTTAA TCACTAAATA TCTTTGCCTA	1620
40	TAGGACTCCA TTGAATACAT TAGCCATTGA TAATCTACCT GTTTAAATGG CCCCTGTTG	1680
	AACTCTCAAG CTTTGAAGAC CTACCTGTTT TTCCAGAAGA GAACGTTGAA AGTGCCATGT	1740
45	TTCCCTTTGC GTGATCTCTG TTGATGGCAC TCTGGAATTG TTTTCAATTAA GTCATTTTAG	1800
	ACATAGCATT TATTATCACT GTGGATCTCT ACTTGTGGG TGTATGAAT TCTTTGAAGA	1860
	AATATATTTT GAAGAGGTGT GGGAGGAAGG AATACATTTT ATAAATGTT GTAGTGAAGC	1920
50	CCACAATTGA CCTTTGACTA ATAGGAGTTT TAAGTATGTT AAAATCTAT ACTGGACAGT	1980
	TACAAGAAAT TACCGGAGAA AAGCTTGTGA GCTCACCAA CAAGGATTTT AGTGTAGATT	2040
55	TTGTCTTTCT TGAACCTAAA GAAACAAATG ACAAAGTTT AATGAAAAG CCTGCTGTTG	2100
	TTCCACATCT CGTTGCTGTT TACATTCCTT TGTGGAGCCT ACATCTTCCT AAGCTTTTTA	2160
	GCAGGTATAT GTTGAACACT TCTGTTTCAT GGTGAGACA GAATCAGAGG CCATGGATAC	2220
60	TGACAACTGA TTTGTCTGTT TTTTCTCTCT GTCTTTTCC ATGACTCTTA TATACTGCCT	2280



CATCTTGATT TATAAGCAAA ACCTGGAAAA CCTACAAAAT AAGTGTGTG GTTATCTAG 2340  
 5 AAAAATATGG AAAATATTGC TGTATTTTTT GGTGAAGAAA ATCAATTTTG TATAGTTTAT 2400  
 TTCAATCTAA ATAAATGTG AATTTTGT TT AAAGCTTAGG CACATTATTT TTGTGGGGT 2460  
 CAAAACATTC TTGTGTAAAT TCTCTTAAAC ATTTGATAAA CAGCTTCACA ATTC 2514

10

(2) INFORMATION FOR SEQ ID NO: 256:

15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2357 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 256:

CTGCCCTTATG AAGCCGATGC AGAAATTTTG GCTGTGAAAT TTCACACTAT GATAACTGAG 60  
 25 AAGTGGGGAT TAAATATGGA GTATTGTCGT GGCCAGGCTT ACATTGTCTC TAGTGGATTT 120  
 TCTTCCAAAA TGAAAGTTGT TGCTTCTAGA CTTTITAGAGA AATATCCCCA AGCTATCTAC 180  
 30 ACACCTCTGCT CTTCCTGTGC CTTAAATATG TGGTTGGCAA AATCAGTACC TGTATGGGA 240  
 GTATCTGTTG CATTAGGAAC AATTGAGGAA GTTTGTCTTT TTTCCATCG ATCACCACAA 300  
 CTGCTTTTAG AACTTGACAA CGTAATTYCT GTTCTTTTTC AGAACAGTAA AGAAAGGGGT 360  
 35 AAAGAACTGA AGGAAATCTG CCATTCTCAG TGGACAGGCA GGCATGATGC TTTTGAAATT 420  
 TTAGTGAAC TCCTGCAAGC ACTGTGTTTA TGTITAGATG GTATAAATAG TGACACAAAT 480  
 40 ATTAGATGGA ATAACTATAT AGCTGGCCGA GCATTGTAC TCTGCAGTGC AGTGTGAGAT 540  
 TTTGATTICA TTGTTACTAT TGTGTCTCTT AAAAATGTCC TATCTTTTAC AAGAGCCTTT 600  
 GGGAAAAACC TCCAGGGGCA AACCTCTGAT GTCTTCTTTG CGGCCGGTAG CTTGACTGCA 660  
 45 GTACTGCATT CACTCAACGA AGTGANTGGA AAATATTGAA GTTATCATG AATTTTGGTT 720  
 TGAGGAAGCC ACAAATTTGG CAACCAAACT TGATATTCAA ATGAAACTCC CTGGGAAATT 780  
 50 CCGCAGAGCT CACCAGGGTA ACTTGGAATC TCAGCTAACC TCTGAGAGTT ACTATAAAGA 840  
 AACCCTAAGT GTCCCAACAG TGGAGCATAT TATTGAGGAA CTAAAGATA TATTCTCAGA 900  
 ACAGCACCTC AAAGCTCTTA AATGCTTATC TCTGGTACCC TCAGTCATGG GACAACTCAA 960  
 55 ATTCAATACG TCGGAGGAAC ACCATGCTGA CATGTATAGA AGTGACTTAC CCAATCCTGA 1020  
 CACGCTGTCA GCTGAGCTTC ATGTGTGGAG AATCAAATGG AAACACAGGG GGAAAGATAT 1080  
 60 AGAGCTTCCG TCCACCATCT ATGAAGCCCT CCACCTGCCT GACATCAAGT TTTTCTCTAA 1140

	TGIGTATGCA TTGCTGAAGG TCCTGTGTAT TCCTCCTGTG ATGAAGGTTG AGAATGAGCG	1200
	GTATGAAAAT GGACGAAAGC GTCTTAAAGC ATATTTGAGG AACACTTTGA CAGACCAAAG	1260
5	GTCAAGTAAC TTGGCTTTGC TTAACATAAA TTTTGATATA AAACACGACC TGGATTTAAT	1320
	GGTGGACACA TATATTAAAC TCTATACAAG TAAGTCAGAG CTTCTACAG ATAATCCGA	1380
10	AACTGTGGAA AATACCTAAG AGACTTTTAA AAATAGGCTT TCTTATATTT GATATTTGGA	1440
	AGAAAAAGCC GTAAGTGTAT GTAGACCACT TAATCACTAA ATATCTTTGC CTATAGGACT	1500
	CCATTGAATA CATTAGCCAT TGATAATCTA CCTGTTTAAA TGGCCCCGTG TTGAACTCTC	1560
15	AAGCTTTGAA GACCTACCTG TTCTTCCAGA AGAGAACGTT GAAAGTGCCA TGTTCCTTTT	1620
	TGCGTGATCT CTGTTGATGG CACTCTGGAA TTGTTTCAGT TAAGTCATTT TAGACATAGC	1680
20	ATTTATTATC ACTGTGGATC TCTACTTGTT GGGTGTATG AATTCTTTGA AGAAATATAT	1740
	TTTGAAGAGG TGTGGGAGGA AGGAATACAT TTTATAAAAT GTTGTAGTGA AGCCCACAAT	1800
	TGACCTTTGA CTAATAGGAG TTTTAAGTAT GTTAAAAATC TATACTGGAC AGTTACAAGA	1860
25	AATTACCGGA GAAAAGCTTG TGAGCTCACC AAACAAGGAT TTCAGTGTAG ATTTTGTCTT	1920
	TCTTGAACCT AAAGAAACAA ATGACAAAGT TTGAATGGAA AAGCCTGCTG TTGTTCCACA	1980
30	TCTCGTTGCT GTTTACATTC CTTTGTGGAG CCTACATCTT CCTAAGCTTT TTAGCAGGTA	2040
	TATGTTGAAC ACTTCTGTTT CATGGTTGAG ACAGAATCAG AGGCCATGGA TACTGACAAC	2100
	TGATTGTCTCT GTTTTTTTTC TCTGTCTTTT TCCATGACTC TTATATACTG CCTCATCTTG	2160
35	ATTTATAAGC AAAACCTGGA AAACCTACAA AATAAGTGTT GTGGTTTATC TAGAAAAATA	2220
	TGGAATAATAT TGCTGTTATT TTTGGTGAAG AAAATCAATT TTGTATAGTT TATTTCATC	2280
40	TAAATAAAAT GTGAATTTTG TTTAAAGCTT AGGCACATTA TTTTTGTGG GGTCAAACA	2340
	TTCTTGTGTA AATTCTC	2357

45

(2) INFORMATION FOR SEQ ID NO: 257:

(i) SEQUENCE CHARACTERISTICS:

50

- (A) LENGTH: 689 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 257:

55

ACTTTCTGGT GCAAAAAGAT GTTCAAGCCT TATTTTATAC TTGCCTGCCC CTTTCTCTTT	60
CATTTATTGG AGTGAGCTGC AGCTCTAAGA AGACCTGTTT TTTTGAATGG AGAGTAGCAT	120
CAGGAACCAG GATGTGGGTG CGAGGCGTGC TCCTGGCTGT TGCAGATTGC TGCACCCGGG	180

60

AGCTCTTAGT GGACAGAGCT AGAGGATATG TGCACGTACT TCCATCTCTC TCTCTGTCTC 240  
CGATTTTAGC CCAGCAACCAC AGGGTACGTT CCAGTTTTC TCTCTTTCCA TAGCTGTAAG 300  
5 GCCCTTTCTG GGAATGGTTC TCATCTCCT TAATCTATTA TTGGGTCAGT TTTCTGTCAT 360  
GTCCCCAGCC TCCCATCACT GCCACCCACT CCCACAGAG ATGCCCTGCT CATCCGACTG 420  
10 GGGCTTTGAC TCCCACACTG TGTACCCCTC TTGTGTGGAC GCCCTGCTGC CAAAACCTTC 480  
AGCAAAACAGC TTTCCAAATG GAAGTTGTCA CTGTCARGGS CTTTACAATC AGCAACAGCA 540  
AAATCTACAT GCTGCTGAGG GTCTGCCTC ATTAAGATGC AATAAATATG TAAGTACATA 600  
15 AAAACAGCAA TAGAAGAAAC GTAATGCTTT ATTCTCAAAT ATGNATGTCT ACATAGAAAA 660  
GCCAAAATTA TTAAGAATAG TAAGGAATT 689

20

(2) INFORMATION FOR SEQ ID NO: 258:

25

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2377 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 258:

TCGACCCACG CGTCCGCCGA TGTGATGATT CCTGCGTATT CCAAGAACCG GGCCTATGCC 60  
35 ATCTTCTTCA TAGTCTTCAC TGTGATAGGG GACGCCCCCG GCGCTGTGCT ATCCTGTGCC 120  
GGCCACCCCT GCGTTGGTTT TGCTGCTGTA CTGGTGGCGC CCCTGACCGT GGCTGTCTCC 180  
TCTTGAAGGA AGCCTGTTTC TGAATGAACCT GCTGACAGCC ATCATCTACA GTCAGTTCCG 240  
40 GGGTACCTG ATGAAATCTC TCCAGACCTC GCTGTTTCGG AGGCGGGTGG GAACCCGGCT 300  
GCCTTTGAAG TCCTATCCTC CATGGTGGGG GAGGGAGGAG CCTTCCCTCA GGCAGTTGGG 360  
45 GTGAAGCCCC AGAATTGCT GCAGGTGCTT CAGAAGGTCC AGCTGGACAG CTCCCACAGA 420  
CAGGCCATGA TGGAGAAGGT GCGTTCCTAT GGCAGTGTTT TGCTCTCAGC TGAGGAGTTT 480  
CAGAAGCTCT TCAACGAGCT TGACAGAAGT GTGGTTAAAG AGCACCCGCC GAGGCCCGAG 540  
50 TACCAGTCTC CGTTTCTGCA GAGCGNCCCA GTTCCTCTTC GGCCACTNAC TACTTTGACT 600  
ACCTGGGGAA CCTCATGCCC CTGGCAAACC TGGTGTCCAT TTGCGTGTTC CTGGTGCTGG 660  
55 ATGCAGATGT TGCTGCCTGC TGAGCGTGAT GACTTCATCC TGGGGGTCT CAACTGCGTC 720  
TTCATGTGT ACTACCTGTT GGAGATGCTG GCTCAAGGTC TTTTGCCCTG GGGCCTGCCA 780  
60 RGGTACYKKT CCTAACCCCA RCAAMGTGTT TTGAACGGGC TCCTCAMCGT TTGTCTGGC 840

	TGGWWKKGSM GATCTCAACT CTGGCTGTGT ACCGATTGCC ACACCCAGGC TGGAGGCCGG	900
	ANATGGTGGG CCTGCTGTCT CTGTGGGACA TGACCCGCAT ACTGAACATG CTCATCGTGT	960
5	TCCGCTTCCT GCGTATCATC CCCAGCATGA AGCCGATGGC CGTGGTGGCC AGTACCGTCC	1020
	TGGGCCTGGT GCAAAACATG CGTGCCTTTG GCGGGATCCT GGTGGTGGTC TACTACGTAT	1080
10	TTGCCATCAT TGGGATCAAC TTGTTTAGAG GCGTCATTGT GGCTCTTCCT GGAAACAGCA	1140
	GCCTGGCCCC TGCCAATAGG TCGGCGCCCT GTGGGAGCTT CGAGCAGCTG GAGTACTGGG	1200
	CCAACAACIT CGATGACTTT GCGGCTGCCC TGGTCACTCT GTGGAACITG ATGGTGGTGA	1260
15	ACAACCTGGCA GGTGTTTCTG GATGCATATC GCGCTACTA AGGCCCTGGG TCCAAGATCT	1320
	ATTTTGTATT GTGGTGGCTG GTGTCTGTCT TCATCTGGGT CAACCTGTTT CTGGCCCTGA	1380
20	TTCTGGAGAA CTTCCTTCAC AAGTGGGACC CCCGAGCCA CCTGCAGCCC CTTCCTGGGA	1440
	CCCCAGAGGC CACCTACCAG ATGACTGTGG AGCTCCTGTT CAGGGATATT CTGGAGGAGC	1500
	CCGGGGAGGA TGAGCTCACA GAGAGGCTGA GCCAGCACCC GCACCTGTGG CTGTGCAGGT	1560
25	GACGTCCGGG TCTGCCATCC CAGCAGGGGC GGCAGGAGAG AGAGGCTGGC ATAACACAGG	1620
	TGCCCATCAT GGAAGAGGCG GCCATGCTGT GGCCAGCCAG GCAGGAAGAG ACCTTTCCTC	1680
30	TGACGGACCA CTAAGCTGGG GACAGGAACC AAGTCCTTTG CGTGTGGCCC AACAACCATT	1740
	TACAGAACAG CTGCTGGTGC TTCAGGGAGG CGCCGTGCCC TCCGCTTTCT TTTATAGCTG	1800
	CTTCAGTGAG AATTCCCTTG TCGACTCCAC AGGGACCTTT CAGACAAAAA TGCAAGAAGC	1860
35	AGCGGCCTCC CCTGTCCCCT GCAGCTTCCG TGGTGCCTTT GCTGCCGGCA GCCCTTGGGG	1920
	ACCACAGGCC TGACCAAGGC CTGCACAGGT TAACCGTCAG ACTTCCGGG CATTCAGCTG	1980
40	GGAATGATAC TAATACCTCC GATTTTAGCC CAGCACCACA GGTACGTTT CAGTTTTTAT	2040
	TTCTTTCCAT AGCTGTAAGG CCCTTTCTGG GAATGGTTAT CATTCCTCTT AATCTATTAT	2100
	TGGGTCAGTT TTCTGTCATG TCCCCAGCCT CCCATCACTG CCACCCACTC CCCACAGAGA	2160
45	TGCCCTGCTC ATCCGACTGG GGCCTTGACT CCCACACTGT GTACCCCTCT TGTGTGGACG	2220
	CCCTGCTGCC AAAACCTTCA GCAAACAGCT TTCCAAATGG AAGTTGTCAC TGTCAGGGCC	2280
50	TTTACAATCA GCAACAGCAA AATCTACATG CTGCTGAGGG TCCTGCCTCA TTAAGATGCA	2340
	ATAAATATGT AAGTACATAA AAAAAAAAAA AAAAAA	2377

55

(2) INFORMATION FOR SEQ ID NO: 259:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1193 base pairs

(B) TYPE: nucleic acid

60

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 259:

5 TCTGNTCGCC GTCGCCCCGC CCCTGGCCTT TGCCCGGTCG GCGGGGACTT CCTGTGTCGT 60  
ATTTCCAAGG ACTCCAAAGC GAGGCCGGGG ACTGAAGGTG TGGGTGTGGA GCCCTCTGGC 120  
10 AGAGGGTTAA CCTGGGTCAA ATGCACGGAT TCTCACCTCG TACAGTTACG CTCTCCGCG 180  
GCAAGTCCGC GAGGMYTTGA AGTCCTGAGC GCTCAAGTTT GTCCGTAGTC GAGAGAAGGC 240  
CATGGAGGTG CCGCCACCGG CACCGCGGAG CTTTCTCTGT AGAGCATTGT GCCTATTTCC 300  
15 COGAGTCTTT GCTGCCGAAG CTGTGACTGC CGATTGCGAA GTCCTTGAGG AGCGTCAGAA 360  
GCGGCTTCCC TACGTCCCAG AGCCCTATTA CCCGGAATCT GGATGGGACC GCCTCCGGGA 420  
20 GCTGTTTGGC AAAGACACAG TGAACACTAG TCTGAATGTA TACCGAAATA AAGATGCCTT 480  
AAGCCATTTT GTAATTGCAG GAGCTGTCAC GGAAGTCTT TTTAGGATAA ACGTAGGCCT 540  
GCGTGGCTGG TGGCTGGTGG CATAATTGGA GCCTTGCTGG GCACTCCTGT AGGAGGCCTG 600  
25 CTGATGGCAT TTCAGAAGTA CTCTGGTGAG ACTGTTTCAGG AAAGAAAACA GAAGGATCGA 660  
AAGGCACTCC ATGAGCTAAA ACTGGAAGAG TGGAAAGGCA GACTACAAGT TACTGAGCAC 720  
30 CTCCCTGAGA AAATTGAAAG TAGTTTACAG GAAGATGAAC CTGAGAATGA TGCTAAGAAA 780  
ATTGAAGCAC TGCTAAACCT TCCTAGAAAC CCTTCAGTAA TAGATAAACA AGACAAGGAC 840  
TGAAAGTGCT CTGAAC TTGA AACTCACTGG AGAGCTGAAG GGAGCTGCCA TGTCCGATGA 900  
35 ATGCCAACAG ACAGGCCACT CTTTGGTCAG CCGCTGACA AATTAAAGTG CTGGTACCTG 960  
TGGTGGCAGT GGCTTGCTCT TGTCTTTTTC TTTTCTTTT AACTAAGAAT GGGGCTGTG 1020  
40 TACTCTCACT TTACTTATCC TTAAATTTAA ATACATACTT ATGTTTGTAT TAATCTATCA 1080  
ATATATGCAT ACATGAATAT ATCCACCCAC CTAGATTTTA AGCAGTAAAT AAAACATTTT 1140  
GCAAAAGATT AAAGTTGAAT TTTACAGTTA AAAAAAAAAA AAAAAAAAAA AAA 1193  
45

(2) INFORMATION FOR SEQ ID NO: 260:

50

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1262 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 260:

60

GAAAAACCCA AAGATGCAGA CAATCTCTTT GAACATGAAT TGGGGGCTCT CAATATGGCT 60

	GCATTACTAC GAAAAGAAGA AAGAGCAAGT CTTCTTAGTA ATCTTGCCCC ATGTTGTAAG	120
	GCGTTGTGCT TCAGACGGGA TTCTGCAATT CGAAAGCAGC TTGTTAAAAA TGAGAAGGGC	180
5	ACCATAAAAC AAGCTTACAC GAGTSCCTCA ATGGTAGACA ATGAATTACT TCGATTGAGT	240
	CTTCGGTTAT TTAAGCGGAA GACTACTTGC CATGCTCCAG GACATGAAAA GACTGAAGAT	300
10	AATAAACTTT CACAGTCCAG TATCCAACAG GAACTGTGTG TGTCTTAAGA CCGAAGTTCA	360
	ATATGGTATT TTTGGTACTG TCTTCCTTCA GCAGTGCATA TTCCTTTGCA AAGTTCTTTG	420
	GTTTGACAAG CATTAGTGAC AAAGGCAGAA AAGATTTATC AGCCATGCTA AAAGAGTGAA	480
15	GAATTTTGAT CTTTAGAGAC ACTAGTTTTG GCCAACTTAA GATTTTACGT TAAITTTTAC	540
	ATAGTATTTG ACACTCATGC AAAATAATGT GAAAACATCT AGATTTAGTA GTTTATTCTG	600
20	CGCCTTTTGT TAAAACTGAA GATTTTGGAA AATGGTTGTC ACTGCTCTTC CAGCCTATGA	660
	ATATTTTTGT GAAATGGAAC CATGGATTTA TGTCTGGATC ATCCATACAG AACCAACAAT	720
	TTTATTCAA AACAATGTGT TCATCAAAGT AATTGCTCAC ATTGTGCAGT ACTATGTTGT	780
25	ACAGACCACG TGAAAGGGAA TGCTGGTCTA GCTGGCGTGG TATGTTTATA GGCGAATTTC	840
	AGCAGAAGGA AGCCAAAATA GTTTTTTCCT TTTGAAAGTT TTTTAAAAAT TATTTTCATG	900
30	GTCTTTTTTT TAATTAATAT GTGTGCATTG TTACAATGTA TGTGGGATGT CTTTGGACCC	960
	TAAATGCTTT TTTTGTATC AGAGATTGTG TACTATTTTT ATTTTAAATA AATGTATCTT	1020
	CCCTTTCCTT GTTTTAGATT TACTTTGCTC TTCGTTAATC TTATTCCTGA TGATCTAGAA	1080
35	CATTAGTCAT CAACATTACA TGTTCATGC TTCAGATATT TTACTGCTTG TGTCTTATT	1140
	GTTGGACAGC TTAAACAGA GTTGATGGTA CTTCAAATAT AGCTCATTGA TACTTAAGGG	1200
40	CANCTTCCTT GGGATGTGGG CTTTTTGGAA GGAAAAAAT TNCCTTAAAG GCAATCCCA	1260
	GT	1262

45

(2) INFORMATION FOR SEQ ID NO: 261:

## (i) SEQUENCE CHARACTERISTICS:

50

- (A) LENGTH: 1179 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 261:

55

GGCAAACCTT CCCCCAANGC TTCGAACTT GCAAGCCGAA ACCTTGAATC GTTAAAGTT	60
GGGTTGCGNC GCGCCCTGG CCCGAAGAAG CGCAATTGGC GTTCCGCGAA CGTTGGCCCT	120
60 CAACGGCTCG GCAGCCAGCC ATGTCCTGCA CCCAGGACAG CGGCCCTGGG CTACAAGGAC	180

	CTGGACCTCA TCTTCCTGCG CCGACCTGCG CGGGGAAGGG GAGTTTCAGA CTGTGAAGGA	240
5	CGTCGTGCTG GACTGCCTGT TGGACTTCTT ACCCGAGGGG GTGAACAAAG AGAAGATCAC	300
	ACCACTCAGC CTCAAGGAAG CTTATGTGCA GAAAATGGTT AAAGTGTGCA ATGACTCTGA	360
	CCGATGGAGT CTTATATCCC TGTCAAACAA CAGTGGCAAA AATGTGGAAC TGAAATTTGT	420
10	GGATTCCTTC CGGAGGCAGT TTGAATTCAG TGTAGATTCT TTTCAAATCA AATTAGACTC	480
	TCTTCTGCTC TTTTATGAAT GTTCAGAGAA CCCAATGACT GAGACATTTC ACCCCACAAT	540
15	AATCGGGGAG AGCGTCTATG GCGATTTCOA GGAAGCCTTT GATCACCTTT GTAACAAGAT	600
	CATTGCCACC AGGAACCCAG AGGAAATCCG AGGGGGAGGC CTGCTTAAGT ACTGCAACCT	660
	CTTGGTGAGG GGCTTTAGGC CCGCCTCTGA TGAAATCAAG ACCCTTCAAA GGTATATGTG	720
20	TTCCAGGTTT TTCATGACT TCTCAGACAT TGGAGAGCAG CAGAGAAAAC TGGAGTCCTA	780
	TTTGACAGAA CACTTTGTGG GATTGGAAGA CCGCAAGTAT GAGTATCTCA TGACCCCTCA	840
25	TGGAGTGGTA AATGAGAGCA CAGTGTGCCT GATGGGACAT GAAAGAAGAC AGACTTTAAA	900
	CCTTATCACC ATGCTGGCTA TCCGGGTGTT AGCTGACCAA AATGTCATTG CTAATGTGGC	960
	TAATGTCACT TGCTATTACC AGCCAGCCCC CTATGTAGCA GATGCCAACT TTAGCAATTA	1020
30	CTACATTGCA CAGGTTACAG CAGTATTCAC GTGCCAGCAA CAGACCTACT CCACTTGGCT	1080
	ACCCTGCAAT TAAGAATCAT TTAAAAATGT CCTGTGGGGA AGCCATTTC AACAAGACAG	1140
35	GAGAGAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAGAGC	1179

40 (2) INFORMATION FOR SEQ ID NO: 262:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1162 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 262:

50	GGCAAACCTT CCCCCAANGC TTCGAAACTT GCAAGCCGAA ACCTTGAATC GTTAAAAGTT	60
	GGGTGTGCGC GGCGCCCTGG CCGAAGAAG CGCAATTGGC GTTCCGCGAA CGTTGGCCCT	120
	CAACGGCTCG GCAGCCAGCC ATGTCCTGCA CCCAGGACAG CGGCCCTGGG CTACAAGGAC	180
55	CTGGACCTCA TCTTCCTGCG CCGACCTGCG CGGGGAAGGG GAGTTTCAGA CTGTGAAGGA	240
	CGTCGTGCTG GACTGCCTGT TGGACTTCTT ACCCGAGGGG GTGAACAAAG AGAAGATCAC	300
60	ACCACTCAGC CTCAAGGAAG CTTATGTGCA GAAAATGGTT AAAGTGTGCA ATGACTCTGA	360

CCGATGGAGT CTTATATCCC TGTCAAACAA CAGTGGCAAA AATGTGGAAC TGAAATTTGT 420  
 GGATTCCTC CGGAGGCAGT TTGAATTCAG TGTAGATTCT TTTCAAATCA AATTAGACTC 480  
 5 TCTTCTGCTC TTTTATGAAT GTTCAGAGAA CCCAATGACT GAGACATTTC ACCCCACAAT 540  
 AATCGGGGAG AGCGTCTATG GCGATTTCCA GGAAGCCTTT GATCACCTTT GTAACAAGAT 600  
 10 CATTGCCACC AGGAACCCAG AGGAAATCCG AGGGGGAGGC CTGCTTAAGT ACTGCAACCT 660  
 CTTGGTGAGG GGCTTTAGGC CCGCCTCTGA TGAAATCAAG ACCCTTCAAA GGTATATGTG 720  
 TTCCAGGTTT TTCATCGACT TCTCAGACAT TGGAGAGCAG CAGAGAAAAC TGGAGTCCTA 780  
 15 TTTGCAGAAC CACTTTGTGG GATTGGAAGA CCGCAAGTAT GAGTATCTCA TGACCCTTCA 840  
 TGGAGTGGTA AATGAGAGCA CAGTGTGCCT GATGGGACAT GAAAGAAGAC AGACTTTAAA 900  
 CCTTATCACC ATGCTGGCTA TCCGGGTGTT AGCTGACCAA AATGTCATTC CTAATGTGGC 960  
 20 TAATGTCACT TGCTATTACC AGCCAGCCCC CTATGTAGCA GATGCCAACT TTAGCAATTA 1020  
 CTACATTGCA CAGGTTGAGC CAGTATTACG GTGCCAGCAA CAGACCTACT CCACTTGGCT 1080  
 25 ACCCTGCAAT TAAGAATCAT TTAAAAATGT CCTGTGGGGA AGCCATTTC AACAAGACAG 1140  
 GAGAGAAAAA NAANGAAAAG AG 1162

30

(2) INFORMATION FOR SEQ ID NO: 263:

35 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 735 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 263:

CCGGCTGGGT ATTTGCCTCG CACCATGCGG CCCAAGGGCA AAGTGGGCAC GAGAGGGAAG 60  
 AAGCAGATAT TTGAAGAGAA CAGAGAGACT CTGAAGTTCT ACCTGCGGAT CATACTGGGG 120  
 45 GCCAATGCCA TTACTGCCT TGTGACGTG GTCTCTTTT ACTCATCTGC CTCATTTTGG 180  
 GCCTGGTTGG CCTTGGGCTT TAGTCTGGCA GTGTATGGG CCAGCTACCA CTCTATGACC 240  
 50 TCGATGGCAC GAGCAGCGTT CTTCTGAGGA TGGGGCCCTG ATGGATGGTG GCACGAGCTC 300  
 AACATGGAGC AGGGCATGGC AGAGCACCTT AAGGATGTGA TCCTACTGAC AGCCATCGTG 360  
 CAGGTGCTCA GCTGCTTCTC TCTCTATGTC TGGTCTTCT GGCTTCTGGC TCCAGGCCGG 420  
 55 GCCCTTTACC TCCTGTGGGT GAATGTGCTG GGCCCTGGT TCACTGCAGA CAGTGGCACC 480  
 CCAGCACCAG AGCACAATGA GAAACGGCAG CGCCGACAG AGCGGCGGCA GATGAAGCGG 540  
 60 TTATAGCCAT TGACATTGTG GCCACAGGCC ACTGGCCCTG GGTGGCTCTG TCAGGGTGCA 600



5 CAGCCCCCTCA TGCCTGGAGC AATGAGGGTC TAGTCCAGGG GCCAAAAGCA GTCTGAGGTA 660  
TTGGGTATAC TTATACTCTA TAGGGTCGTT GAATAAATGG CTTAGAATGT GAAAAAAAAA 720  
AAAAAAAAAA ATTTT 735

10

(2) INFORMATION FOR SEQ ID NO: 264:

15 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 783 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 264:

AAGTGCATGA GCTGCCGATG TGGTGCTTAG TGATTGCGGT TTCGGTCGCT CTCCCGTGTT 60  
TCCCGGGCTG GGTATTTGCC TCGCACCATG GCGCCCAAGG GCAAAGTGG CACGAGAGGG 120  
25 AAGAAGCAGA TATTTGAAGA GAACAGAGAG ACTCTGAAGT TCTACCTGCG GATCATACTG 180  
GGGGCCAATG CCATTTACTG CCTGTGACG TTGGTCTTCT TTTACTCATC TGCCTCATTT 240  
TGGGCTGGT TGGCCTGGC TTTAGTCTGG CAGTGTATGG GGCCAGCTAC CACTCTATGA 300  
30 GCTCGATGGC ACGAGCAGCG TTCTCTGAGG ATGGGGCCCT GATGGATGGT GGCATGGACC 360  
TCAACATGGA GCAGGGCATG GCAGAGTGAG TGTCCTCCAC CGCCAGCCCA GGCACCTTAA 420  
35 GGATGTGATC CTA CTGACAG CCATCGTGCA GGTGCTCAGC TGCTTCTCTC TCTATGTCTG 480  
GTCCTTCTGG CTCTGGCTC CAGGCCGGC CCTTTACCTC CTGTGGGTGA ATGTGCTGG 540  
CCCCTGGTTC ACTGCAGACA GTGGCACCCC AGCACCAGAG CACAATGAGA AACGGCAGCG 600  
40 CCGACAGGAG CGCGGCAGA TGAAGCGGT ATAGCCATTG ACGATTTKGC SACNRGCCAC 660  
TGGCCCTGGG TGGCTCTGTC AGGGTGACA GCCCCTCATG CCTGGAGCAA TGAGGGTCTA 720  
45 GTCCAGGGGC CAAAAGCAGT CTGAGGTATT GGGTATACTT ATACTCTATA GGGTCGTGA 780  
ATA 783

50

(2) INFORMATION FOR SEQ ID NO: 265:

55 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1638 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

60 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 265:

	GGCACGAGGC GCGGCAGCG GTGGCGGCGG CGCCCCCGG CGGGAGCCGT NCCCTTTCCC	60
5	GTGGGGGAGC GCGGGYCGG GGYCCAGGG ANCCCGGMC ACGGAGAGCG GGAAGAGGAT	120
	GGATTGCCCG GCCTCCCCC CCGGATGGAA GAAGGAGGAA GTGATCCGAA AATCTGGGCT	180
	AAGTGTCTGC AAGAGCGATG TCTACTACTT CAGTCCAAGT GGTAAAGT TCAGAAGCAA	240
10	GCCTCAGTTG GCAAGGTACC TGGGAAATAC TGTGTATCTC AGCAGTTTIG ACTTCAGAAC	300
	TGAAAGATG ATGCCTAGTA AATTACAGAA GAACAAACAG AGACTGCGAA ACGATCCTCT	360
15	CAATCAAAAT AAGGGTAAAC CAGACTTGAA TACAACATTG CCAATTAGAC AAACAGCATC	420
	AATTTTCAA CAACCGTAA CCAAAGTCAC AAATCATCCT AGTAATAAAG TGAAATCAGA	480
	CCCACAACGA ATGAATGAAC AGCCACGTCA GCTTTTCTGG GAGAAGAGGC TACAAGGACT	540
20	TAGTGCATCA GATGTAACAG AACAAATTAT AAAAACCATG GAACTACCCA AAGGTCTTCA	600
	AGGAGTTGGT CCAGGTAGCA ATGATGAGAC CCTTTTATCT GCTGTTGCCA GTGCTTTGCA	660
25	CACAAGCTCT GCGCCAATCA CAGGGCAAGT CTCCGCTGCT GTGAAAAGA ACCCTGCTGT	720
	TTGGCTTAAC ACATCTCAAC CCCTCTGCAA AGCTTTTATT GTCACAGATG AAGACATCAG	780
	GAAACAGGAA GAGCGAGTAC AGCAAGTACG CAAGAAATTG GAAGAAGCAC TGATGGCAGA	840
30	CATCTTGTG CGAGCTGCTG ATACAGAAGA GATGGATATT GAAATGGACA GTGGAGATGA	900
	AGCCTAAGAA TATGATCAGG TAACTTTTGA CGACTTTCC CCAAGAGAAA ATTCTTAGAA	960
35	ATTGAACAAA AATGTTTCCA CTGGCTTTTG CCTGTAAGAA AAAAAATGTA CCCGAGCACA	1020
	TAGAGCTTTT TAATAGCACT AACCAATGCC TTTTATAGT TATTTTGTAT GTATATATCT	1080
	ATTATTCAAA AAATCATGTT TATTTTGAGT CCTAGGACTT AAAATTAGTC TTTTGTAATA	1140
40	TCAAGCAGGA CCTAAGATG AAGCTGAGCT TTTGATGCCA GGTGCAATCT ACTGGAAATG	1200
	TAGCACTTAC GTAAACATT TGTTCCCCC ACAGTTTAA TAAGAACAGA TCAGGAATTC	1260
45	TAAATAAATT TCCAGTTAA AGATTATGT GACTTCACTG TATATAAACA TATTTTATA	1320
	CTTTATTGAA AGGGGACACC TGTACATTCT TCCATCRTCA CTGTAAAGAC AAATAAATGA	1380
	TTATATTAC AGACTGATTG GAATTCCTTC TGTGAAAAG CACACACAAT AAAGAACCCC	1440
50	TCGTTAGCCT TCCTCTGATT TACATTCAAC TCTGATCCCG GGCCTTAGG TTTGACATGG	1500
	GAGGTGGGAG GAAGATAGCG CATATATTG CAGTATGAAC TATTGCCTCT GGGACGTTGT	1560
55	GAGGAATTGT GCTTTCACCA GAATTTCTAA GGATTTCTGG CTAAATATC ACCTAGCCTG	1620
	TGGTAATTTT TTTTCCCT	1638

## (2) INFORMATION FOR SEQ ID NO: 266:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1455 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 266:

5  
 10  
 15  
 20  
 25  
 30  
 35  
 40  
 45  
 50  
 55  
 60

CGTGGCTACT GCCATGCAGG TACCGGGTCC GGAATTCCCA GGGTCGACCC ACCCGTCCGC 60  
 TCAGTTGGCA AGGTACCTGG GAAATACTGT TGATCTCAGC AGTTTGTACT TCAGAACTGG 120  
 AAAGATGATG CCTAGTAAAT TACAGAAGAA CAAACAGAGA CTGCGAAACG ATCCTCTCAA 180  
 TCAAAATAAG GGTAACCAG ACTTGAATAC AACATTGCCA ATTAGACAAA CAGCATCAAT 240  
 TTTCACAAAC CCGTAACCA AAGTCACAAA TCATCCTAGT AATAAAGTGA AATCAGACCC 300  
 ACAACGAATG AATGAACAGC CACGTCAGCT TTTCTGGGAG AAGAGGCTAC AAGGACTTAG 360  
 TGCATCAGAT GTAACAGAAC AAATTATAAA AACCATGGAA CTACCCAAAG GTCTTCAAGG 420  
 AGTGGTCCA GGTAGCAATG ATGAGACCCT TTTATCTGCT GTTGCCAGTG CTTTGACAC 480  
 AAGCTCTGCG CCAATCACAG GGCAAGTCTC CGCTGCTGTG GAAAAGAACC CTGCTGTTTG 540  
 GCTTAACACA TCTCAACCCC TCTGCAAAGC TTTTATTGTC ACAGATGAAG ACATCAGGAA 600  
 ACAGGAAGAG CGAGTACAGC AAGTACGCAA GAAATTGGAA GAAGCACTGA TGGCAGACAT 660  
 CTGTGCGGA GCTGCTGATA CAGAAGAGAT GGATATTGAA ATGGACAGTG GAGATGAAGC 720  
 CTAAGAATAT GATCAGGTAA CTTTCGACCG ACTTTCCCA AGAGAAAATT CCTAGAAATT 780  
 GAACAAAAT GTTCCACTG GCTTTGCTT GTAAGAAAAA AAATGTACCC GAGCACATAG 840  
 AGCTTTTAA TAGCACTAAC CAATGCCTTT TTAGATGTAT TTTGATGTA TATATCTATT 900  
 ATTCAAAAA TCATGTTTAT TTTAGTCTT AGGACTTAAA ATTAGTCTTT TGTAAATATCA 960  
 AGCAGGACCC TAAGATGAAG CTGAGCTTTT GATGCCAGGT GCAATCTACT GGAAATGTAG 1020  
 CACTTACGTA AAACATTTGT TTCCCCACA GTTTTAATAA GAACAGATCA GGAATTCTAA 1080  
 ATAAATTTC CAGTTAAAGA TTATTGTGAC TTCCTGTAT ATAAACATAT TTTTATACTT 1140  
 TATTGAAAGG GGACACCTGT ACATTCTTCC ATCCTCACTG TAAAGACAAA TAAATGATTA 1200  
 TATTACAGA CTGATTGGAA TTCTTTCTGT TGAAAAGCAC ACACAATAAA GAACCCCTCG 1260  
 TTAGCCTTCC TCTGATTTAC ATTCAACTCT GATCCCGGG CCTTAGGTTT GACATGGGAG 1320  
 GTGGAGGAA GATAGCGCAT ATATTGTCAG TATGAATAT TGCTCTGGG ACGTTGTGAG 1380  
 GAATTGTGCT TTCACCAGAA TTCTAAGGA TTTCTGGCTT AAATATCACC TAGCCTGTGG 1440  
 TAATTTTTTT TCCCT 1455

## (2) INFORMATION FOR SEQ ID NO: 267:

5

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1086 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

10

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 267:

15

CGCCTGCAGT ACCGGTCCGG AATTCCCGGG TCGACCCACG CGTCGCTGAC CCAGGAGAAG 60

CTGCCCTGTCT ACATCAGCCT GGGCTGCAGC GCGCTGCCGC CGCGGGGCGG GCAGCTGAAC 120

TATGTGCTCT TCAGGGCGGG CACCGTGTG CATTCATCTT TGTACCCCA GCATCTAGCA 180

20

GTGTTGGCAT GTAGTAGGCA CTCAAGAAAT GTGTGTGAA TGAACGATGC CTGTGACAAG 240

CAAGCGGACT TTATTCTTTC CTGACCCTTG CTCCTATGAC ACACCTCCTC CTGACTGCCA 300

25

CTGTCACTCC TTCAGAGCAG AACTCCTCTA GGGAACTGG ATGGGAAACA GCCATGGCCA 360

AGGACATCCT GGTGAAGCA GGGTACACT TTGATGAACT GAACAAGCTG AGGGTGTGG 420

ACCCAGAGGT TACCCAGCAG ACCATAGAGC TGAAGGAAGA GTGCAAAGAC TTTGTGGACA 480

30

AAATTTGCCA GTTTCAGAAA ATAGTTGGTG GTTTAATTGA GCTTGTGAT CAACTTGCAA 540

AAGAAGCAGA AAATGAAAAG ATGAAGGCCA TCGGTGCTCG GAACTTGCTC AAATCTATAG 600

35

CAAAGCAGAG AGAAGCTCAA CAGCAGCAAC TTCAAGCCCT AATAGCAGAA AAGAAAATGC 660

AGCTAGAAAG GTATCGGGTT GAATATGAAG CTTTGTGTAA AGTAGAAGCA GAACAAAATG 720

AATTTATTGA CCAATTTATT TTTTCAGAAAT GAACTGAAAA TTTGCTTTT ATAGTAGGAA 780

40

GGCAAAACAA AAAAAAGCCT CTCAAAACCA AAAAAACCTC TGTAGCATTC CAGCGGCTTG 840

ACCAATGACC TATGTCACAA GAGTGCGGT GTAAGGAATG CAGCCCCCTG AAGACAGCAC 900

45

TACAAGTCTG GGGGAGCCAG TTTTAACATC AGTGCACAGC TGCTGCTGGT GGCCCTGCAG 960

TGTACGTTCT CACCTCTTAT GCTTAGTTGG AACTAAGCAG TTTGTAACT TTCATCCTTT 1020

TTTTTGTAAT TTCACAAAGC TTTGAAGGA GARGCAATAA ATTTTGTGTT TCNAAATGGC 1080

50

TTGATG 1086

55

## (2) INFORMATION FOR SEQ ID NO: 268:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1003 base pairs

(B) TYPE: nucleic acid

60

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 268:

5 GGCACGGGAG CAGCCGGGCT GGTCTGCTG CGAGCCGGCG GCCCGGAGTG GGGCGGCGGA 60  
 GCAAACATGA ACGTTGGAGT TGCCACAGT GAAGTGAATC CAAATACCG TGTCTGAAC 120  
 AGCCGGGGTA TGTGGCTGAC ATATGCATG GGAGTTGGCT TGCTTCATAT TGTCTTACTC 180  
 10 AGCATTCCTT TCTTCAGTGT TCCTGTTGCT TGGACTTTAA CAAATATTAT ACATAATCTG 240  
 GGGATGTACG TATTTTTGCA TGCAGTGAAA GGAACACCTT TCGAACTCC TGACCAGGGT 300  
 15 AAAAGCAAGG CTCCTAACTC ATTGGGAACA ACTGGACTAT GGAGTACAGT TTACATCTTC 360  
 ACGGAAGTTT TTCACAATTT CTCCAATAAT TCTATATTTT CTGGCAAGTT TCTATACGAA 420  
 GTATGATCCA ACTCACTTCA TCCTAAACAC AGCTTCTCTC CTGAGTGTAC TAATTCCTCA 480  
 20 AATGCCACAA CTACATGGT TCCGATCTT TGGAAATTAAT AAGTATTGAA ATGTTTTGAA 540  
 ACTGAAAAAA AATTTTACAG CTAAGAATT TCTTATAAGG AAGGAGTGGT TAGTAACTG 600  
 25 CACTGTTTCT CTGATAATGT GAAATGAGAA GTATTTACAT TGGAGGGCCA ATGGCTGGTC 660  
 CTTCAAGTGC TGTTTTGAAG TGCAGATTTC CATTAATGA TGCCTCTGTT TAATACACCT 720  
 GGTACATTTC TGAAGAGGGG CTTTATAAGC AGGCTGGGCA GGGCCAGCTT ATAAGTTAAA 780  
 30 GGGCATCACA GTGAGGGTGT AGTAGATAAA TTCAAGGAAA TAAGAGATTT GTAAGAACT 840  
 AGGACCAGCT TAACTTATAA TGAATGGGCA TTGTGTTAAG AAAAGAACAT TTCCAGTCAT 900  
 35 TCAGCTGTGG TTATTTAAAG CAGACTTACA TGTAACCGG AATCCTCTCT ATACAAGTTT 960  
 ATTAAAGATT ATTTTATTA CCGTAAAAAA AAAAAAAAAA AAA 1003

40

(2) INFORMATION FOR SEQ ID NO: 269:

(i) SEQUENCE CHARACTERISTICS:

- 45 (A) LENGTH: 1234 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 269:

ATCAGCATCT ACAAGTAGCA TATTTTGGAT GGTGTTGTG TGCTACTTCA AAGTAACTAG 60  
 GAAAAAATAA TCCTCGCAAC ACAGGTACCT TGTCTGTCA GAATGGGGG TGTAGGTTG 120  
 55 CCAGTGTAT CAGTGTGAT TCATTTCAAT ACTTCCTACA GAGCAAACAT GAACGTTGGA 180  
 GTTGCCACA GTGAAGTGAA TCCAAATACC CGTGTCTGA ACAGCCGGG TATGTGGCTG 240  
 60 ACATATGCAT TGGGAGTTGG CTGCTTCAT ATGTCTTAC TCAGCATCC CTTCTCAGT 300

	GTTCCTGTG CTTGGACTTT AACAAATATT ATACATAATC TGGGGATGTA CGTATTTTGTG	360
5	CATGCAGTGA AAGGAACACC TTTCGAAACT CCTGACCAGG GTAAAGCAAG GCTCCTAACT	420
	CATTGGGAAC AACTGGACTA TGGAGTACAG TTTACATCTT CACGGAAGTT TTTCACAATT	480
	TCTCCAATAA TTCTATATTT TCTGGCAAGT TTCTATACGA AGTATGATCC AACTCACTTC	540
10	ATCCTAAACA CAGCTTCTCT CCTGAGTGTG CTAATTCCCA AAATGCCACA ACTACATGGT	600
	GTTCGGATCT TTGGAATTAA TAAGTATTGA AATGTTTTGA AACTGAAAAA AAATTTTACA	660
15	GCTACTGAAT TTCTTATAAG GAAGGAGTGG TTAGTAAACT GCACTGTTTC TSTGATAATG	720
	TGAAATGAGA AGTATTTTACA TTGGAGGGCC AATGGCTGGT CCTTCAAGTG CTGTTTTGAA	780
	GTGCAGATTT CCATTAAATG ATGCCCTCTGT TTAATACACC TGGTACATTT CTGAAGAGGG	840
20	GCTTTATAAG CARGCTGGGC AGGCCAGCT TATAAGTTAA AGGCCATCAC AGTGAGGGTG	900
	TAGTAGATAA ATTCAAGGAA ATAAGAGATT TGTAAGAAAC TAGGACCAGC TTAACITATA	960
25	ATGAATGGGC ATTGTGTAA GAAAAGAACA TTTCCAGTCA TTCAGCTGTG GTTATTTAAA	1020
	GCAGACTTAC ATGTAAACCG GAATCCTCTC TATACAAGTT TATTAAAGAT TATTTTATT	1080
	ACCRTACATA TTTCKCTTGT TTTATGTAAG YGGATGTATA TCCTCTTGT TTATACAAGC	1140
30	CAGTCCAC TTATGAGGT ACTTTTTTGG TTTTGCTGGG CTTAATATG TGTATTGGTC	1200
	AATGAGGCCA TTTTACANT TATTAACGTT ACAG	1234

35

(2) INFORMATION FOR SEQ ID NO: 270:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 574 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 270:

	NGAGGTGCGT TCTGAGCCGT CTGTCCTGCG CCAAGATGCT TCAAAGTATT ATTAAAAACA	60
50	TATGGATCCC CATGAAGCCC TACTACACCA AAGTTTACCA GGAGATTTGG ATAGGAATGG	120
	GGCTGATGGG CTTTCATCGTT TATAAAATCC GGGCTGCTGA TAAAAGAAGT AAGGCTTTGA	180
	AAGCTTCAGC GCCTGCTCCT GGTCACTACT AACCAGATTT ACTTGGAGTA CATGTGAAAG	240
55	AAAACGTCAG TCTGCCTGTA AATTTTCAGCA AGCCGTGTTA GATGGGGAGC GTGGAACGTC	300
	ACTGTACACT TGTATAAGTA CCGTTTACTT CATGGCATGA ATAAATGGAT CTGTGAGATG	360
60	CACTGCTACC TGGTACTGCT TTCAGTGTGT TCCCCCTCAG CCCTCCGGCG TGTCAGGCAT	420

ACTCTGAGTA GATAATTTGT CATGCAGCGC ATGCAATCAG AATCTCACTG AGCCACCCAT 480  
 CATGTGAAA TAATTACCTC AGTTGTACAG GACTTGGTGA TCAGGATCCA GGCACCTCACT 540  
 5 TGTATTCTAC TGCTCAATAA ACGTTTATTA AACT 574

10 (2) INFORMATION FOR SEQ ID NO: 271:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1731 base pairs  
 (B) TYPE: nucleic acid  
 15 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 271:

20 GCTGCAAGGT GCGCCTCGTG CCGCTGCAGA TCCAGCTCAC TACCCTGGGA AATCTTACAC 60  
 CTTCAAGCAC TGTGTTTTTC TGCTGTGATA TGCAGGAAAG GTTCAGACCA GCCATCAAGT 120  
 25 ATTTTGGGGA TATTATTAGC GTGGGACAGA GATTGTTGCA AGGGGCCCCG ATTTTAGGAA 180  
 TTCTGTAT TGTAAACAGAA CAATACCCTA AAGGTCTTGG GAGCACGGTT CAAGAAATTG 240  
 ATTTAACAGG TGTAAGAACTG GTACTTCCAA AGACCAAGTT TTCAATGGTA TTACCAGAAG 300  
 30 TAGAAGCGGC ATTAGCAGAG ATTCCCGGAG TCAGGAGTGT TGTATTATTT GGAGTAGAAA 360  
 CTCATGTGTG CATCCAACAA ACTGCCCTGG AGCTAGTTGG CCGAGGAGTC GAGGTTTACA 420  
 TTGTGTGCTGA TGCCACCTCA TCAAGAAGCA TGATGGACAG GATGTTTGCC CTCGAGCGTC 480  
 35 TCGCTCRARC CNGGGATCAT AGTGACCACG AGTGNAGGCT GTTCTGCTTC AGCTGGTAGC 540  
 TGATAAGGAC CATCCAAAAT TCAAGGAAAT TCAGAATCTA ATTAAGGCGA GTGCTCCAGA 600  
 40 GTCGGGTCTG CTTTCCAAAG TATAGGACAT TTGAAGAACT GGTATGCTAC TCACTGGTGA 660  
 AGGACAGTCA GGTGAAGGAC TGTAAAGCCA CACAAGCTCT TCTTATCTCT ACTAGAATTA 720  
 AAATGTTAAG TCAAAAACGG CTCCTTTTTT GCGCCTCCTA GTGAACCTAA CCAGCTAGAC 780  
 45 CATTTGAGTA CCAGCATTTA GTTACAAACG TCAAAGGCTT CCGGTGCTGC TTACCTTCCT 840  
 TTTTGTATA TGTGCTTTTA TTTATTAAAA AAAATTACAA TGAAGATGCC TGTTTTGTCT 900  
 50 CTACTGTGTA CTCGTATCGT ATCTTTCCAA AGTGCAGACT CTTGTGAAGT TTTCTTAAAT 960  
 TGTTCACTTT AAAGAAAATG ACGTACCAAC AATGATTTGG CTTTTATATT ACTGTAAGAT 1020  
 55 GTTATAATGT TAATGTGGAT GTAGTGCTTT TACTTTACAG ATTGATTGGA ATAAGATTAT 1080  
 TGCATATGAA TTTACCCACA GGACTCTGAA TCATGTTACC CACTCCCTC ACAATGTTGT 1140  
 CCACTTAGTG AGTTGCATTG ATCTATCCGT ACCAAATGAT GTTGAATAAT TACATATCTT 1200  
 60 TCTKGACTAT ACTGATTTCT TATTTTGGTC ACTATTACTA AATCTCTGTT AATAATCTCT 1260

5 CTTTAACTG AAAAGGGATG GGATAGAAGG GTTGTCAATG CCATATTATT GGTGGAGGGC 1320  
 TGTTTTAACA TCCTTGAAGT ATGGCTTGCT GAATATCTTT ACCAACATCT TGAATATATA 1380  
 TTCTAGTGTC CACAAGATTT AGCAAAAAGA TAAAGCTTGG GTGGAATATC ATTTTAAAT 1440  
 GTTCATGPTC TGTCTATAT TTTCTTCACC TACTCTCCAA ATATTGTAAT GCAAAAAGTC 1500  
 10 TCAGTAATGA TTTGGTAGTA TTAATTTTGT GGTCAATTGT TCTCTTCGAT AAATTTATTT 1560  
 TCATTAAATA CTTRITAGAG GGTTTTGAAA TGTPTTTTCAA ATATGTGAAA TGTGAAACTG 1620  
 CTGTCTTTTA TATTAAAGTA ATTAAAGAAA ATGTATTGTG ATTGAAATTA TTTTGNCTC 1680  
 15 CACAAGATGG CTCTATGAGT ATTCTTCCAG GGATTCTAAT ATTTATTTAA G 1731

20

(2) INFORMATION FOR SEQ ID NO: 272:

25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1320 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 272:

CTGCTTAGGA AGAGAAGGTC AGAGTTCGCG GGGGCAGAGG CATTCTTGCC GCTGGCCAG 60  
 TCACTATGTA GTGGAGGGGC AGACACCCTC CCGCAAATTC TGAAGGTTT TTAGTCTCGA 120  
 35 CTAGGGCAGT AGCCAGGAC TCCTAGTCGC CGGCTTCAGG TCACTGCCCG CTGAACGGAG 180  
 CTGCCGTCGC CATGTTTGGC TGCTTGGTGG CGGGGAGGCT GGTGCAAACA GCTGCACAGC 240  
 AAGTGGCAGA GGATAAATTT GTTTTGGACT TACCTGATTA TGAAAGTATC AACCATGTTG 300  
 40 TGGTTTATAT GCTGGAACA ATCCCATTTT CTGAGGGAAT GGGAGGATCT GTCTACTTTT 360  
 CTTATCCTGA TTCAAATGGA ATGCCAGTAT GGMAACTCCT AGGATTGTGTC ACGAATGGGA 420  
 45 AGCCAAGTGC CATCTTCAAA ATTTCAAGTC TTAAATCTGG AGAAGGAAGC CAACATCCTT 480  
 TTGGAGCCAT GAATATTGTC CGAACTCCAT CTGTTGCTCA GATTGGAATT TCAGTGGAAAT 540  
 TATTAGACAG TATGGCTCAG CAGACTCCTG TAGGTAATGC TGCTGTATCC TCAGTTGACT 600  
 50 CATTCACTCA GTTCACACAA AAGATGTTGG ACAATTTCTA CAATTTTGCT TCATCATTTG 660  
 CTGTCTCTCA GGCCAGATG ACACCAAGCC CATCTGAAAT GTTCATTCCG GCAAAATGTG 720  
 55 TTCTGCAAAT GGTATGAGGC ATMTCTGTG TCCAATATTA AGGCTTTTTA TAACTGAATA 780  
 TCTATTTTGT CTATGAATAT ATTCTTTTTT TGACATTTAA ACATATCTTT TTATTGTGAA 840  
 60 CATCAGCACT GCATGCCATT AAAGTATGTA CTATAGAGAT CTGATGAGAA ACAGTTCTTA 900



CCCTAAATAT TTTGTTATAT TGTCGCCATT ATGAATTTAT AAAGACAGGA AAATATAGTT 960  
GCCTATGTTT TAGGGACCAC TATTAAAGCT TATAAATATT TGTGTATTTT CATTTAGAAG 1020  
5 TACCATCTAT GAGAGTAGTT TATACTGCAC TGTGTACATG AATGGCTAAT GAATCTATTT 1080  
TCCAACCTTC CCGTGTTTTA TAGATATTTT TTTTCACTTT GAGTATCCTA GAGATGGGAG 1140  
10 GATGCCTAGG AAGAGTTTGT TGAGAAGTGG TACCATGGTG TAGCATGGGA GAGCATTGGG 1200  
AATGCACTAG GTTTGAATTT GGCATAATGG TAGCTATGTG ACCCTGAGCA AATTTCTCTC 1260  
ATCTGCTCAT CTGANGAATG AGGAAATAGG AGTGAATTTG ATNTTTCCTA GGTCCNICTA 1320  
15

## (2) INFORMATION FOR SEQ ID NO: 273:

- 20 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 515 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 273:

CCCTGGAGAG GGGCTGCTGT GCCAGCTTGG GGAGGCTCTG GGATGGGGCT GCCCCTGATG 60  
30 GCCCTGATGT GGAGTACCTT GCCAGCATCT GCTGGGGTGA ACTTTATTTT AGCCCTTCCC 120  
TTGTTGYTCT TATGAAGAAC AGAGGAGGGG TGGGCAGGTC AGTGATGTCA GCAGTGAGTA 180  
35 TTCCCAGCAC AGCGGCTCTG GAAGAGGCAT GAGGCATTTT TTTTCAGGAA TGRTCATTAT 240  
TCAGCCAGAA GGCATTTCATT AAGTAAGTCC TGACTTTGTG CCCAGCTCTG TGTATAGGC 300  
CCTTGGCGAG ACTCAGGAGG GGCARAGGAC GCTAGKTTKT AGWTAACACG GAACCTCARA 360  
40 GGWTATATGG TCCAAGAAGA CCCGGGGGCG GTGAAAACCC TGTGGACTAA TGCTCACGGG 420  
AGCCCCAGGT CACACTTTGA CTTTGCTACC ATGGGCTGTG TCTANGNACG TATATATGCT 480  
45 GCGTAATTAT TACAGAGGCA GTCCATGTGC ATTGT 515

## (2) INFORMATION FOR SEQ ID NO: 274:

- 50 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2995 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 274:

60 TGACACCCAT AAGGAATTCA TGAAGAAAGT AGAAGAAAAG CGAGTGGACG TTAACTCAGC 60

	AGTAGCCATG GGAGAAGTCA TCCTGGCTGT CTGCCACCCC GATTGCATCA CAACCATCAA	120
	ACACTGGATC ACCATCATCC GAGCTCGCTT CGAGGAGGTC CTGACATGGG CTAAGCAGCA	180
5	CCAGCAGCGT CTTGAAACGG CCTTGTGAGA ACTGGTGGCT AATGCTGAGC TCCTGGAAGA	240
	ACTTCTGGCA TGGATCCAGT GGGCTGAGAC CACCCTCATT CAGCGGGATC AGGAGCCAAT	300
10	CCCGCAGAAC ATTGACCGAG TTAAAGCCCT TATCGCTGAG CATCAGACAT TTATGGAGGA	360
	GATGACTCGC AAACAGCCTG ACGTGGACCG GGTACCAAG ACATACAAA GGAAAAACAT	420
	AGAGCCTACT CACGCGCCTT TCATAGAGAA ATCCCGCAGC GGAGGCAGGA AATCCCTAAG	480
15	TCAGCCAACC CCTCCTCCCA TGCCAATCCT TTCACAGTCT GAAGCAAAAA ACCCAGGAT	540
	CAACCAGCTT TCTGCCCGCT GGCAGCAGGT GTGGCTGTTA GCACTGGAGC GGCAAAGGAA	600
20	ACTGAATGAT GCCTTGGATC GGCTGGAGGA GTTGAAAGAA TTTGCCAACT TTGACTTTGA	660
	TGTCTGGAGG AAAAAGTATA TGCCTTGGAT GAATCACAAA AAGTCTCGAG TGATGGATTT	720
	CTTCCGGCGC ATTGATAAGG ACCAGGATGG GAAGATAACA CGTCAGGAGT TTATCGATGG	780
25	CATTTTAGCA TCCAAGTCC CCACCACCAA GTTAGAGATG ACTGCTGTGG CTGACATTTT	840
	CGACCGAGAT GGGGATGGTT ACATTGATTA TTATGAATTT GTGGCTGCTC TTCATCCCAA	900
30	CAAGGATGCG TATCGACCAA CAACCGATGC AGATAAAATC GAAGATGAGG TTACAAGACA	960
	AGTGGCTCAG TGCAAAATGTG CAAAAAGGTT TCAGGTGGAG CAGATCGAG AGAATAAATA	1020
	CCGGTCTTC CTGGCAATC AGTTTGGGA TTCTCAGCAG TTCCGGCTGG TCCGTATTCT	1080
35	GCGCAACCGT GATGGTTCCG GTTGGTGGAG GATGGATGGC CTGGATGAA TTTTTAGTGA	1140
	AAAATGATCC CTGCCGAGCA CGAGGTAGAA CTAACATTGA ACTTAGAGAG AAATTCATCC	1200
40	TACCAGAGGG AGCATCCAG GGAATGACCC CCTTCCGCTC ACGGGGTCGA AGGTCCAAAC	1260
	CATCTTCCCG GGCAGCTTCC CCTACTCGTT CCAGCTCCAG TGCTAGTCAG AGTAACCACA	1320
	GCTGTACATC CATGCCATCT TCTCCAGCCA CCCCAGCCAG TGAACCAAG GTTATCCCAT	1380
45	CATCAGGTAG CAAGTTGAAA CGACCAACAC CAACTTTTCA TTCTAGTCGG ACATCCCTTG	1440
	CTGGTGATAC CAGCAATTAG TTCTTCCCG GCCTCCACAG GTGCCAAAAC TAATCGGGCA	1500
50	GACCCTAAAA AGTCTGCCAG TCGCCCTGGG AGTCGGGCTG GGAGTCGAGC CGGGAGTCGA	1560
	GCCAGCAGCC GCGAGGAAG TGACGCTTCT GACTTTGACC TCTTAGAGAC GCATTGCTTG	1620
	TTCCGACACT TCAGAAAGCA GCGCTGCAGG GGGCCAAGGC AACTCCAGGA GAGGGCTAAA	1680
55	CAAACCTTCC AAAATCCCAA CCATGTCTAA GAAGACCACC ACTGCCTCCC CCAGGACTCC	1740
	AGGTCCCAAG CGATAACACT GTCTAAGCAC CCCCAGCCA CTATCCACTT TGAATCCTGC	1800
60	TCCATACATT GGGTGTATAT TTATTCTGAA CGGAGAAGT TATATTGTTA AAAGTGTAAG	1860

	AGAATAATTG TGTATGAAG CTGCCTTATT TTTTCTCTT TTGTAAGTGA CTATTTTCAT	1920
	GTGAATATTT ATGTAGATAA AATTTCCTC CTGTAACCC TGTATGGAT GGGGCCAGA	1980
5	AATGAAATAT TTGAGAAAA CAAGTAAAA GGTCAAGATA CAAATGTGTA TTAACAAAAA	2040
	AAAAGCCTAT TAATAGGGTT TCTGCGCGT GCAGGGTGT AAACCTGCTT TATCTTTAG	2100
10	GATTATTCCT AAATGCATCT TCTTATAAA CTGACTTGC TATCTCAGCA AGATAAATTA	2160
	TATTAAAAA ATAAGAATCC TGCAGTGTTC AAGGAACCTT TTTTGTGTA ATCACGGACA	2220
	CCTCAATTAG CAAGAACTGA GGGAGGGCT TTTTCCATTG TTTAATGTTT TGTGATTTT	2280
15	AGCTAAAGAG AGGGAACCTC ATCTAAGTAA CATTTGCACA TGGATACAGC AAAAGGAGTT	2340
	CATTGCAATA CTGCTTTGG ATATTGTTTC AGTACTGGGT GTTTAAAGGA CAAATAGCTG	2400
	CTAGAATTC AAGGTAAATG TAAGTGTTC AAAAAAGTCA GAACATTTGG GGTTTTAAAC	2460
20	TGATTTGTTG CTCCCTATCC AGCCTAGACA CCAGTAACTC TTGTGTTTAC CAGGACCCAG	2520
	ACCCITGGCA AGGATAGGC TCGTTGGTGA CATTGTGAAT TTCAGATTG TTTTATCCAC	2580
25	TTTTTTTGTCT ATTTATTTAA ATGGTCGATC AACTTCCAC AAAGTAGGA ATGAATTCCA	2640
	CGAGCCTGTT CTGAAAATGT GGACGTAAGA CAAACACGTG CTCGTCTTTT AATGGAGTTC	2700
30	ACCAGCACAC TTGTTAACCA GTCTGTTTG CTTCGTCTT TTTTGTGCG TAATAAAGTC	2760
	AACTGACCAA GTGACCATGA AAAGGGGCTG TCTGGGGCTC CTGTTTTTTA GCTGCTGTTT	2820
	TTCAGCTCCG ACCATGTTGC TGTGTGATTA TCTCAATTGG TTTAATTTGA GGCAGAACT	2880
35	GAAGCTCTAC CAATGAACTG TTTAGAAACA AGACACACTT TTGTATTAAA ATTGCTTGCA	2940
	GTAACAAAAA AAAAAAAAAA AAAAAAAAAA AAAAACTCG AGGGGGGCCC GGTAC	2995

40

(2) INFORMATION FOR SEQ ID NO: 275:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1990 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 275:

	GGGACCCGCG CGSCTCCCGG GGATGGTGAG CAAGGCGCTG CTGCNWCCTG TCTGCCGTCA	60
	ACCGCAGAGG ATGAAGCTGC TGCTGGGCAT CGCCTTGCTG GCCTACGTG CCTCTGTTTG	120
55	GGGCAACTTC GTTAATATGA GGTCTATCCA GGAAAATGGT GAACTAAAAA TTGAAAGCAA	180
	GATTGAAGAG ATGGTTGAAC CACTAAGAGA GAAATCAGA GATTAGAAA AAAGCTTTAC	240
60	CCAGAAATAC CCACCAGTAA AGTTTTTATC AGAAAAGGAT CGGAAAAGAA TTTTGAWTAA	300